

# Bringing Order to Chaos: ScienceBase and Other Project Lifecycle Tools

## **Speakers:**

Emily Fort, Data and Information Coordinator  
USGS National Climate Change and Wildlife Science Center

## **Recording Date:**

Thursday, September 5, 2013

**Ashley Isham:** Good afternoon from the U.S. Fish and Wildlife Service's National Conservation Training Center in Shepherdstown, West Virginia. My name is Ashley Isham, and I would like to welcome you to today's broadcast of the NCCWSC Climate Change Science and Management Webinar Series. This series is held in partnership with the U.S. Geological Survey's National Climate Change and Wildlife Science Center in Reston, Virginia.

Today's speaker, Emily Fort, will be presenting "Bringing Order to Chaos: ScienceBase and Other Project Lifecycle Tools."

I'm joined by Dr. Shawn Carter, who will now introduce Emily. Shawn, welcome.

**Shawn Carter:** Thank you and welcome everyone. Today it's my pleasure to introduce Emily Fort, who is the Data and Information Coordinator here at the National Climate Change and Wildlife Science Center in Reston, Virginia.

At USGS, Emily's responsible for coordinating IT and data management activities among the Climate Science Centers, the CSCs, and also our center, including developing data policies, and leading an Information Management Working Group. Before coming to USGS, Emily was a program manager for federal collaboration initiatives at OMB.

Today, she is going to be talking about "Bringing Order to Chaos: ScienceBase and Other Project Lifecycle Tools." It's my pleasure to introduce Emily.

**Emily Fort:** Thanks, Shawn. I'm here today to talk about what the Climate Science Centers and the National Climate Change and Wildlife Science Center have been doing in the arena of data management and what tools we're using, which include ScienceBase and several other things.

This grew out of a talk that Shawn Carter gave a few months ago, where he talked about the science and what the National Climate Change and Wildlife Science Center is all about and mentioned some of the data things that we were working on.

It got a lot of questions, so I'm here to talk a little bit more. This covers a lot of different topics. If folks have any follow-up, I'm happy to chat. Any questions, we'll just step right through it.

First, a little bit of an agenda. I'm just going to do a quick briefing of who we are and some of our challenges and opportunities. Talk about how we've approached this issue, some of the policies we've put in place, a bit about our process, and the big picture. Dive into that big picture a little bit. Talk about some integration things that we've been working on and where we see things heading.

To summarize a little bit of the initial -- when I came on board a few years ago, we wanted to figure out how we could best support the National Climate Change and Wildlife Science Center, the CSCs' data management, and provide some core tools that would support the activities that were ongoing.

Just in case folks aren't familiar with our enterprise, I just wanted to do a really quick overview of the Climate Science Centers and the National Climate Change and Wildlife Science Center. There are eight CSCs spread regionally throughout the U.S. They're partnerships with institutions, with universities.

In most cases, you'll see that there's more than one university in each region that supports the CSC in addition to the national office here in the DC area. We have this unique nature where we're working with both federal organizations and universities to fund science.

Some of the challenges that we face from a data perspective -- some of these are really basic, obviously -- is we needed to know what we were doing and where across the network.

I always go back to the very basic. If my boss is going to brief someone on the Hill or at OMB, he may get a question or need to know what are we doing in the area of birds or fish or in this particular geographic region, and that information should be easily accessible.

We needed to provide some basic capabilities and ensure there was some consistency across the enterprise. We developed some data policies, and we wanted to make sure that those were being followed.

Whatever we did had to provide access for both university and federal scientists. We also recognized that we were funding about a hundred projects every year and that, over time, that number, as it grew, it was going to be really important to have some systems and tools in place, or else it was going to become pretty unmanageable.

We were a new program and growing, with a pretty small staff. We needed to work pretty innovatively to find solutions for some of these issues.

That leads to the opportunities. We did have a pretty blank slate. We were a new program, just started a few years ago. We didn't have many of these tools. Even things like a website, we had to figure out how that was going to happen.

There have been more and more supportive government policies. I'm sure many of you guys are familiar with some of the open data policies coming out of the White House and the executive office that have really enabled some of our activities in the areas of data and data sharing.

In addition, we also have a really supportive management team. I always say that the fact that I'm merely in my position indicates some level of the importance which is placed on data management and data sharing and tools, because they recognize the need for someone to coordinate those activities and to make sure that things were happening there.

We had extensive capabilities. At USGS, at other federal agencies and universities, people were doing a lot of really cool things that we wanted to take advantage of.

A few slides, a few notes on our approach. One was just really basic. Don't build another stove pipe. I didn't see any need to build something specific for us because there were a lot of tools out there that did things similar to what we needed. What could we leverage? What could we add to? What could we build on?

We needed to provide support to the CSCs, both with tools and with people. This isn't going to happen without some help from people making it happen. We needed some policies to provide some teeth to what we were trying to do.

We wanted to be able to link the projects that the Climate Science Centers and the National Center were doing to the data and the products that flowed out of those so we could understand what we were producing, where it came from, and where it might go.

It was really important to use standards and web services so that we could make sure that what we were doing would be as open to integration and reuse and consistent with good scientific practices as possible and, as part of this, identify some opportunities for integration or where we might be able to link tools or spend just a little bit of money to add value.

We also just wanted to focus on some really core capabilities that our partners could link to, integrate, and maybe even re-imagine and do something really cool with what we have started.

Just a quick note starting with the data policies because they're the foundation to what we've done. The first thing we worked on was the data sharing policy, which was based on some other similar federal policies, which basically just said that what you do as a result of the CSC funding, that information will be shared.

I put "unless there's a good reason not to" simply because, in a few cases, for example, endangered species or perhaps some tribal activities, there may be a good reason we don't want to show nitty-gritty data.

But, in general, data should be shared. That sharing needs to happen when the project is complete. No sitting on it for years and years and years and then, when you're close to retirement, saying, "Oh, hey, by the way, here's this hard drive full of data."

We require data management plans for all funded activities, and we identified some common standards that we'd like to be used. Of course, metadata must be provided. That had to be a core part of what we were doing. The link to the policies is right below.

This is a quick schematic, big picture. I'm not going to go into one of these, because I'm going to be talking about it. It gives you a sense of ScienceBase, which I'll dive into in a little bit more detail here in a second. It's at the center, and a lot of things flow in and out and are connected to that.

Then, at the bottom, the data stewards, who are the people. We have one person for every Climate Science Center, who works with the funded researchers and helps with data management. They are really key in the underpinnings to making this whole enterprise work.

First I want to talk a quick second about RFP Manager. This is a tool we just started this past year because we needed a way to collect proposal information. Previously, our directors managed this through email. The feedback was cumbersome, hard to manage. All the information was lost in their emails.

Getting it and rolling it up, or aggregating it, or from my perspective, understanding the data management plans and policies was pretty challenging.

We worked with the team at the Fort Collins Science Center at USGS to take an existing tool that they already had in place and we thought would be a good jumping off point for the RFP Manager. We added some capabilities, and what we could pretty easily -- nothing, nowhere near, as complicated as, for example, NSF's proposal management system.

This allowed us to get some basic information about who was submitting a proposal, have the review conducted through the system, be able to track who was doing what, collect things like a budget and a data management plan, as well as the full proposal document. Then track which ones we were going to provide funding to.

One key aspect of this, at least from my perspective, is the data management plan. We asked each proposal to fill out a data management plan and, as I mentioned, we have a template that's at that site.

What we we've done is organized it by inputs and outputs. Every research project is taking either existing data and/or collecting new data and then doing something with it to produce one or more outputs or products or new datasets.

For each of those inputs and outputs there were just a few key fields that we needed to understand. Not all of these are required when you're submitting a proposal. We really tried to think about what we needed to know to help evaluate the proposal. Then if you were funded, we went back and asked for the full table.

This is some basic metadata about what they're doing -- brief, brief description. What sort of format is it in? What sorts of quality checks have been done? Backup and storage? Approximately how large is it? The couple of key things really were if there are any issues with exclusive use, like on data sharing or restrictions on its reuse.

What we found through going through all of the DMPs from this last cycle is that there were a few cases where people said, "This is data I collected on my own time that I'm using as an input." Knowing that at the beginning allowed us to ask some questions about if that would have any implications on sharing the results of their research or have that negotiation.

The other important aspect of collecting the data management information early on is that it simply sets that expectation. The researchers know this is something we're going to be interested in. When we work with them and when we touch base with them as they move toward completion of their project, they're not, hopefully, surprised to learn that this is something that we're really interested in. We are going to expect them to transition those products and datasets to -- at least a copy -- to the national center for storage.

One thing that we did as part of this whole RFP cycle, the Climate Science Centers coordinated a call for proposals. The data steward team reviewed each and every data management plan. We sat down as a group and provided comments, discussed questions. We learned a lot, discovered some areas where we can continue to improve.

Then we identified some questions, maybe some gaps, or some things that were done really well, as well, and provided those comments back to each Climate Science Center's director. The Climate Science Center directors could work with their PIs to answer any questions and make sure that everyone was comfortable with the plan and our approach moving forward. That was really helpful and really raised our comfort level with what we're funding and what we're working with.

Then at the completion of the project the plan is for the data stewards to work with the research team to transfer their products and their datasets to the NCCWSC repository, which is ScienceBase, which I'll get to in just a second.

Here we are. What is ScienceBase? You guys may have heard a lot about it or you may be really familiar with it. It's a tool that was developed at USGS. It's a website. It provides some data cataloging and a data management platform for USGS science system partners. It has a great search interface. It's an open source project. It's more than just the CSCs. Lots of other projects, lots of other groups, are using it.

Why did we choose ScienceBase? What attracted us to it? I mentioned earlier in the approaches that we didn't want to build another stove pipe. ScienceBase had several criteria that were really appealing. It was very powerful. It could take lots of different types of data. It wasn't built around one type of science, which is important because Climate Science Centers do a lot of different types of science.

We needed to be able to handle everything from biological data to climate models to a vulnerability assessment, and everything else in between and beyond. It had some core fields, but

it was also extensible, so we could add to it additional fields as we needed. We worked with the team to do that if we needed it.

It was very searchable. It had a geospatial component. Another important aspect was it was available and accessible to non-DOI users. It had permission controls. We could have everything open or we could restrict it, if needed.

The other blessing from my perspective is it met all of the federal IT and security requirements. The team at Fort Collins that manages ScienceBase and others at USGS, they get to worry about all of that stuff that, in my opinion, isn't super fun. I just get to work with them to make sure that the tool does what we need it to do, which I really appreciate.

Also to have web services built in, a REST service architecture that enabled a lot of what we wanted to do to happen.

I'm just going to walk through a little bit about different aspects of ScienceBase to give you a little bit more detail. The first is just "finding stuff," because ScienceBase has everything from datasets to publications, to lots of different items as well.

There's a basic search, which is what you see here on the left. You can search by keyword, you can search by tag, you can search geospatially, lots of different ways to get at it. There's an advanced search, which has even more detail.

Another important thing that took us some time to work out, which is how we were going to organize our information. ScienceBase is designed around communities. Each community has the ability to assign its own permissions and organizations. Each community also allows for its own cataloging, managing and sharing of information.

We knew we were organizing around projects. The way we think about things are what fiscal year was that project funded? We had each Climate Science Center, and then, projects funded in a fiscal year. Then below that, we had to think a little about how we wanted to organize that information, so we did it as this folder structure you'll see here on the right hand side, where it says "Approved Datasets," "Approved Products," "BASIS," "Other" and "Working."

What that did was give us a designated area to put datasets, products. BASIS is, for those of you who may not be familiar, a USGS system that USGS project information goes into. For us, it was limited because it was limited to USGS only, but ScienceBase does harvest from BASIS, so what this allows us to do is to take advantage of whatever information might be in BASIS, and try to minimize any duplications, but also to be able to enhance it where we need to. Where there isn't a BASIS record, obviously, there's no need to worry about the linking.

You'll notice that we also have this working area, which gave us a place to put things that weren't quite ready for prime time, maybe some information that was still in the review process or some working documents that we wanted to have stored centrally and organized together, but we didn't want out on the public web.

When you look in ScienceBase and you look at one of our project records, this is an example of what you'll see. Organized in the central section with a description, some information on the PI, the start and end dates, and then, if there's a footprint. You can provide geospatial information in a lot of different ways. You can upload a shape file or a GeoTIFF or even draw a polygon using a tool that's inside ScienceBase. There's a lot of ways to get that geospatial information.

Then, on the left hand side, you'll see that there's an area where you can tag it with whatever keywords that might be useful. On the right hand side, you'll see where it fits in this structure. Then, if for example, in a lot of the CSC cases, the project may be co-funded with an LCC, a Landscape Conservation Cooperative, so we can designate that relationship or have the same record show up in multiple communities, if that's appropriate. In this case, you can see that that relationship does exist.

I'll need to back up one. You'll see here in the upper right this "Manage" button. If I click that, I'm logged in now. I'll be able to edit this record. Obviously, this is where the permissions get really important. There's people that have edit rights, but not anyone can come in and edit this record. You have to have the appropriate rights.

When you do edit it, you'll see the first basic who, what, when, where, how type of questions. Also an area for you to upload files. That extensions tab is where that flexibility really comes into it, because there's a lot of extensions for custom fields. You can have one for citations, if this record is related to a publication. You can have one related to more budget information, if it's related to a project record. You can really customize it and add the information as you need to.

One thing we had to do with this was think about -- since we wanted to be able to present this information on our website -- what fields we were going to use to map to our website, so that everything would show up consistently and appropriately. That took us some time to think through and work through. Of course, there's always a little bit of complexity. Those were some of the initial decisions that we had, but we worked through that.

Then we took advantage of the ScienceBase web services. For those of you that are super techy, I'm sure you're well familiar with these. For those of you that aren't, you can just think about basically this is a way to get information from one website and present it on another website through magic. That's basically what web services allow you to do.

In this case, ScienceBase, again for the techy lovers, it has a REST service architecture, using JSON. For those of you who don't care, you can just ignore that, nod your head and think about web services.

A couple of examples. I'm going to dive into are the NCCWSC website project pages and also a tool called "DEPTH." You'll see those in just a second.

This is a screenshot of the project pages and a URL example is on the left. All of this information is on our website now. It's organized by CSC and fiscal year. It all comes automatically from ScienceBase.

If we go in and add a file to ScienceBase or we add a project record, almost instantaneously, it shows up on our website. We don't have to do anything to maintain this content. Simply by maintaining our project information in ScienceBase and using those web services, we're able to present all this on our website, which is really powerful.

Here you'll see just the list of projects. On the right hand side, you'll see little icons. If it has a folder icon, that means that there's some sort of product, like a publication, a fact sheet, a presentation or a dataset. If it has the map icon, that means that there's some sort of geospatial information. That's the quick summary.

If you click on one of those, you'll go to a detailed project. Again, this is one of our examples of a project that was finished a couple years or so ago. You can see here, we can have a nice map about where the project is, find out some basic information about the project -- who the PIs were, start and end date, the tags, a brief description -- but then, we have all together, all in one place, all the products associated with that project. We have their fact sheet, their report, and data.

Down at the bottom, you'll see data and maps. People can come here, read about the project, read the report, and if the data is useful to them or they can reuse it, they can grab that as well. That's just an example.

Also using web services, we did another tool called "DEPTH." ScienceBase is very flexible and it handles lots of different types of data. While that has a lot of good aspects, it isn't designed around a project focus, so when you looked at that project page, it maybe wasn't presented how you would ideally want it to get the project information. What DEPTH has is strictly filtered and focused on projects.

There's lots of different filters here. You can filter on organization, fiscal year, if it's a completed project or in process, who the PI, keywords, status, the project title. The other thing that we added here is several of our CSCs have really been focusing on trying to do regional coordination, so to understand in their region what other organizations, including other federal agencies or maybe the LCCs were doing and how they complemented each other and how they could work best together.

This really helps promote that because they were able to tag all of their projects with their science agenda, so we know which aspect of their science planning that this project supports. You can search by that and find, where the information exists, projects potentially spanning multiple agencies and really understand what's being done in a particular area related to their science planning.

I've talked a little bit about several different tools. Now, I want to talk about how they're related or how they fit together.

RFP manager, which is the tool that we use to manage the proposals and review the data management plans, is really the first step in the process. What we've done is develop a harvest that collects all the information for the funded proposals in RFP manager allows you with a click of a button to move those into ScienceBase.

That really helps us because last year a poor student that I hired had to sit and go through all of our proposals and re-key all of the records into ScienceBase. It wasn't a ton of fun. It's also prone for errors, so this really enables that to happen more automatically. We can just do a quick review and make sure when it's ready to move to the public.

Once it's in ScienceBase, that allows us to then consume those web services and make the information available on our website, in Depth, or anywhere else people might be using the same services. You can find it via the ScienceBase search or an integrated search tool, which I'm going to talk about here in just a second.

Another question you might be asking yourself is, "OK, how does this work with other systems and other stuff?"

This is a diagram at a high level that talks a little bit about that, where you have lots of different data generations. There's tons of websites out there. There's tons of projects. They are all feeding into ScienceBase, which has a set of capabilities.

Then another tool, which I'll mention here in a second, the Geo Data Portal, which is really spectacular at managing some large data support and some web processing services. It works with a lot of, especially our larger, datasets. Then, you can take information out of these and feed them into what you need them to, whether that be a desktop model you're running, or an analysis you want to do in Data Basin or ArcGIS or however you might want or need to work with that information.

I mentioned the Geo Data Portal. I do want to highlight that. Down at the bottom, you'll see there is a link. They did a webinar a few months ago that was really great. You can go and learn lots more about it. Just to highlight it, it's a tool that really helps scientists use some subsets, some of these really large models and datasets around their particular interest.

For example, you need a climate model. A lot of the popular ones have been made available through the Geo Data Portal and a lot of the ones we funded. Instead of having to pull it down and then figure out how to get the bit just for your research area, Geo Data Portal will help you do that. It's a really great tool that can help speed up the process and manage that information.

It's standards based, which is a love and a theme, and it's integrated with ScienceBase, so they can work pretty nicely together.

Finally, I wanted to talk about the integrated search. This was a prototype that we did with ScienceBase, the Geo Data Portal and a tool at the University of Idaho, called the "Northwest Knowledge Network." They're one of our CSC universities in the Northwest. We identified a common metadata standard and used a web service called the CSW, the Catalog Service for the Web, to allow you to go to one place, which is on our website, and search across all three catalogs at the same time.

You can find information of interest. In this case, the example in this screenshot, I searched on climate. You can find what might exist on that topic across all three. There's an advanced search to enable you to fine tune it even more. We see this as a way to acknowledge that there's a lot of

different places people might go for information and to make it easier for people to find it by giving one common interface to get at that information.

We're also looking for feedback. This is relatively new. We'd love to hear what works, what doesn't work, and what you might like to see added.

I showed this big picture at the beginning. I'm coming back to it because now I've talked about a lot of these items, so you can see RFP Manager flowing into ScienceBase. We're actually developing a Data Management Plan editor that will also flow into ScienceBase and from RFP. How the information from ScienceBase flows into our website and DEPTH, which actually can flow both ways because DEPTH -- I failed to mention, but it also allows you to create records there which gives you a much more intuitive interface for creating project records.

Then on the right hand side are several of the applications that we work with. You've got the Geo Data Portal. I didn't touch on it, but they've also developed a Derivative Portal which is really cool and powerful. And the Integrated Search which is related to Idaho's Northwest Knowledge Network. All this is floating together.

Then lastly at the bottom, we're working on a tool to collect information on Vulnerability Assessments. But we're still thinking about how it might be related to other things, as well.

One question just in case it comes up. I know several of you are from LCCs and you might wonder how we're working together or how things are related. ScienceBase is also known by another name, LC MAP and that's used a lot in the LCC community. That's the same tool, just with slightly different styling and focus. But that enables that things that are developed for us can be reused by them and vice versa.

Also, the ScienceBase team at Fort Collins is part of the LCC network's Integrated Data Management Network project. I'm a part of that as well. That's really looking at taking some of the capabilities that are part of ScienceBase and several of these other tools and linking them together, and thinking about how we might do things better.

Everything that we've done in ScienceBase can be leveraged by other parts of USGS and other partners. Then both the LCC community and the CSC community have a data management working group. There's cross fertilization on both of those groups so we stay in pretty close touch about who's doing what and working together.

So what did all this buy us? Was this worth the time and energy? In my opinion, of course, I think it was.

We provided some, really core, basic capabilities for the CSCs that helped the enterprise have the basics it needed to grow and understand. It's given us a lot more complexity and allowed us to grow in other areas. We've invested in these tools that then people can reuse and benefit from that value and grow. So that's great.

I think we've really got a good foundation for the program's growth and its future. And we've supported the good data management principles that we really believe in and that are the foundation of what we're doing.

So where do we go from here? One, we're constantly learning. We are new and we recognize that the first time, we probably didn't get it perfect. So every time we do a cycle of reviews or work with PIs we learn where we can improve our guidance, clarify things. How we can walk that line between getting the right information from them, especially, when they're submitting proposals to make sure that we're understanding what they're trying to do.

We want to continue to identify areas for future collaboration and integration with partners and CSC members. I'm sure there will be additional tools and features that are needed and we'll want to develop.

Then we're continuing to add some features. Some things that we're working on right now with the ScienceBase team are adding the ability to easily add Digital Object Identifiers for all of our datasets and to improve the metadata tools and capabilities.

Of course, as the government policies evolve and grow, we'll be staying on top of those and aligning with them.

We've reached the end. I think I can take some questions.

But first I want to -- I couldn't do this without a lot of help. The team at Fort Collins, and the team at the Center for Integrated Data Analytics that does the Geo Data Portal and Holly Padgett, here, in my office. We all work closely together and it helps make this happen. They're great and awesome to work with. Also, I should mention the data stewards. I have a great group, and I appreciate them very much.

I'm ready for questions, Ashley.

**Ashley:** All right. Excellent. Thank you very much, Emily. Great presentation. We had excellent attendance. I think, at one point, it was up to 133. So a lot of interest here.

**Pat Watson:** Yeah. I was wondering if you were going in and mining any data to produce metrics on your program?

**Emily:** That's certainly one thing we want to do, Pat. It's on the list. We're still thinking about exactly what some of those metrics might be. But that's certainly something we've talked about and it's on the pathway, as well.

**Pat:** OK. Thanks.

**Ashley:** Then we have a chat question from Glen. It says, "How broadly used is the RFP Manager tool?"

**Emily:** We started it. It was our baby this year. For just the CSC RFP, we got about 400 -- we do it in two phases, the Statement of Interest and then Proposals. We got about 400 Statements of Interest and about 100 Proposals.

There's already been several different groups that have learned about what we're doing. The Fort team is actually working on a more generic version of RFP Manager that could be used, potentially, by other groups.

I know USGS, I believe, the Community for Data Integration has expressed some interest. There have also been some super high level chats with some other groups. I think there will be more and more. It's just simply because it's a pretty nice little tool for small more straightforward proposal cycles.

**Ashley:** OK. Thank you. Then we have another question that says, "Are Geo Data Portal records harvested into ScienceBase or do you need to use the integrated search?"

**Emily:** I'm not actually sure if they are harvested. I don't think they are, but if Dave or anyone from Fort can answer affirmatively -- Ashley, can they hit \*6 to unmute themselves?

**Ashley:** Absolutely.

**Emily:** Feel free to jump in guys.

I guess, I'm going to go with "no." If that is different, I can get back in touch with whoever -- who asked that question?

**Ashley:** Zhahai Stewart.

**Emily:** OK.

**Ashley:** And I...

**Dave Blodgett:** Actually, Emily. Emily this is Dave Blodgett. I do not believe that they are, because there's not an extension in ScienceBase to support the web service interfaces that the Geo Data Portal uses.

**Tim Kern:** This is Tim Kern. Sorry.

**Emily:** Please, Tim, go ahead.

**Tim:** That's correct. The data from Geo Data Portal are not, those records are not harvested by ScienceBase yet until we have that supported.

**Emily:** But that's, always, these things can be worked out. It's just where they fall on the priority list.

**Ashley:** All right. It says, "Can the LCCs individually use the RFP manager for their own RFPs?"

**Emily:** You would have to work with the Fort team and determine level of support and how that was going to work. But technically, absolutely, and I'm sure they'd be more than happy to work with you.

**Ashley:** All right. Thank you very much. We have another one from Josh and it says, "Where is the ScienceBase source code available?"

**Emily:** Tim or Fort team, I'm going to defer to you guys as well for that.

**Tim:** Hi, Josh. This is Tim Kern. There's a Git repository. Because there's a little bit of a configuration involved, we handle that repository. Right now, we allow anybody who wants it can get the source code. We just ask them to register with us.

This is Tim Kern. You can contact me at ScienceBase@usgs.gov and request that.

**Ashley:** Could you just type that to all participants in the chat box? Your email address?

All right. We have one from Peter, now, and it says, "Is this the new USGS data repository for all unpublished journal data that is not in another data system? There was a mention of how these would get a DOI?"

**Emily:** What we've been working on has been really funded and focused on Climate Science Center projects and fundings, and making sure that those items have a home, as I like to say.

I believe that there is a broader look at USGS especially to support some of the open data initiatives and some of those calls to look at, potentially, how they might handle or provide a home for some information that doesn't have it. And also to get DOIs with those activities. But that's beyond what we're doing. I'm not exactly sure on the status of those items.

But, I believe, certainly, if we add, for example, the ability in ScienceBase to get a DOI, I think USGS does have a relationship with an organization to grant those. So once we do that and have that in ScienceBase, it would be one of those items that other groups and other activities using the tool could take advantage of.

**Ashley:** All right. Then from Ron, it says, "Is it possible to manage project time lines, Gantt charts, et cetera?"

**Emily:** I'm assuming that's through ScienceBase. ScienceBase doesn't have that level of project support. You could, of course, upload those as attachments or as files, but it doesn't have built-in those project management tools.

**Ashley:** All right. Then going back to Zhahai's question, it says, "What does it take to get records harvested into ScienceBase from a local database?"

**Emily:** I think there's several different options for that. Again, because that's beyond a little bit how we've been thinking about things. If Tim wants to jump in, I believe that the team has been working on web accessible folders. Also if there are web services, I'm sure they could look at

how to make that happen. That would be, probably, a discussion that you'd have to have. Tim, I'll let you better answer that.

**Tim:** Again, if anyone has specific questions, that ScienceBase@usgs.gov email address will work. But as far as getting your records harvested, ScienceBase has a harvester interface that you set up your site for. You set up your area for harvesting. So it's a matter of just a little bit of training to do that. It's really not bad.

**Ashley:** OK. Thank you very much. Then we do have a question of, "What exactly is behind the sail boat?"

**Emily:** [laughs] I was trying to find something zen and pretty to put on the last slide.

That's a picture that I took on my vacation this summer from the San Juan Islands. Behind that is another San Juan Island. It's that area of the country off the coast of Washington.

**Ashley:** Excellent. Sounds like fun. All right. Are there any more questions?

**Female Participant 1:** Ashley?

**Ashley:** Yes?

**Female Participant 1:** I wanted to ask how you guys are dealing with copyright? Are you putting publications in there?

**Ashley:** Are you referring to ScienceBase?

**Female Participant 1:** Yeah.

**Ashley:** OK. Emily?

**Emily:** Generally, rather than duplicate the content, we'll try to do links whenever we can. So we'll do a link or a citation to the article.

If it's, of course, a federal scientist, some of those things are a little bit more straightforward. If it's a university scientist, of course, we want to acknowledge things like copyright, make sure we're following the rules.

Those are some details that we're still nailing down exactly how best, writing up some best practices. But in general, we do a lot of linking.

**Female Participant 1:** OK. One more question. How do your scientists like it? How do your project chiefs deal with making their inputs?

**Emily:** We provide a lot of support so they don't actually have to go in themselves into ScienceBase and do much. Although, they can if they want to. Generally, they're pretty busy at this point.

I think, in general, what we found working with the scientists, there's always some push back. This is a bit of a new idea for some people. And some people have to get a little comfortable with it. But by and large, actually, most people have been pretty supportive of this effort of making information available and giving their work a home and being able to highlight it.

What's really been powerful, actually, is what we've done with a lot of our funded scientists is go and show them the project pages and say, "Look. This is going to be a place that we're going to highlight the great work that you're doing. Giving you an opportunity to highlight all the products and the work that you've done. You can point to that and link to it and include it as part of your CV and your professional career."

And they've gotten pretty excited. They like that idea of being able to have that and demonstrate what they've accomplished. I think it's a matter of providing those incentives and helping them understand the value and the big picture, as well.

**Female Participant 1:** Sounds pretty efficient. Thank you.

**Ashley:** Excellent. Thank you.

Transcription by CastingWords