>> Thank you, Amy. It is truly my privilege to introduce Dr. David Tarboton. Dr Tarboton is a Professor of Civil and Environmental Engineering in the Utah Water Research Laboratory at Utah State University. He received a doctorate in Civil Engineering from MIT in 1989. And has been on the faculty at Utah State University for 28 years where he teaches Hydrology and Geographic Information Systems and Water Resources. And where his research advances science's capability for hydrologic prediction. By developing models that take advantage of information and process understanding enabled by new technology. He has developed a number of models and software packages include--including the [inaudible] hydrologic terrain analysis and channel network extraction package. And the Utah energy balance [inaudible] model. Most importantly, certainly from OAC's perspective, David has been the prime mover behind HydroShare, a collaborative environment for sharing hydrologic data and models. Operated by the Consortium of Universities for the Advancement of Hydrologic Sciences, Incorporated or CUAHSI. And truly one of the key science cyber infrastructures that [inaudible] has invested in. For those who are new to HydroShare, HydroShare has enabled scientists to freely and openly share their research artifacts, data, models, workflows, and of course publications. Through contributions from developers ranging from hydrology graduate students to seasoned developers. Which has resulted in accelerated and transformative advances in hydrological sciences. Several awards have been forthcoming from his many contributions to hydrological science. He received the 2015 Utah State University Faculty Research of the Year Robbins Award in 2015. In 2016, CUAHSI presented him with a community service award for his work on hydrologic information systems and HydroShare. And most recently, David was named a fellow of the American Geophysical Union. So congratulations, David. And with that, over to you.

>> Thank you very much. It's great to be able to present this afternoon. You should now be seeing my first slides. HydroShare is a web-based system for sharing hydrologic data and models with specific functionality aimed at making collaboration easier for researchers and models. I'll point out that you could actually go to HydroShare and actually see slides right now if you did a search for OAC Webinar as, as the keyword. It's operated, as [inaudible] mentioned by the Consortium of Universities for the Advancement of Hydrologic Science, Incorporated. With ongoing developments from the team that's listed there at Utah State University, BYU, the [inaudible], University of Illinois, Tufts University of Virginia and [inaudible] University of North Carolina which is where the hardware that hosts HydroShare is hosted. And they support the hardware as well as the software engineering aspects of it. And I'm really indebted to the team of people who've worked on it. Some of the names are there. Not the names of everybody. But it's really great to have had, had all their support for this. This is really our, the hydrology community's contribution to supporting the demands for transparency, research reproducibility, and open data that are increasing. And in this webinar, I'll describe what HydroShare is. How it's been developed as cyber infrastructure to address data management and collaboration needs of the hydrology community. And how this has engaged hydrologists, programmers, and software professionals in the development. It'll describe our work to try and increase the value of research investments and data collection through enabling its longevity. And reuse. It's really intended to be be a cyber infrastructure, web-based hydrologic information system that can fit. And interoperate in the sort of larger ecosystem of cyber

infrastructure elements that we've sometimes seen expressed as a vision from, from the, from OAC. And I'm just going to express here that vision as I understand it. And I hope that HydroShare can be part of it. This vision is not fully realized. And it's certainly not all done by HydroShare, but I think we are working towards an ecosystem. With elements that make up an integrative cyber infrastructure for collaboration and computation that will integrate data storage, data organization, discovery analysis and modeling through web applications. That allows researchers to employ services beyond their desktop and make data storage manipulation reliable and scalable. While improving the ability to collaborate. So some of the reasons for wanting to do that is that you can then do work without needing your own software and tech form to use. You can reduce the challenges of installation and library, software library and [inaudible] dependencies. You can work with larger data sets. Reuse data, reproduce the work that you're doing in a transparent way. Thereby enhanced trust in the findings. And support collaboration. And I think it's really collaboration that is driving a lot of what we're doing. So in terms of motivating it, I think it's collaborative research. I like to say that hydrology is a team sport, and I think that applies to many disciplines. Advancing hydrologic understanding. Requires integration of information from multiple sources. Due to diverse types of data and models. And the pictures to the right sort of depict some of them. We have time series data. We have information about stream metrics and watersheds in the form of geographic features. We have multidimensional data depicted in the, in the cube that you see there in [inaudible] time dimensions. You have gridded data. You actually have models that might be represented with a flow chart that's--figure at the bottom there. And working with the data sets may be a data and computation intensive and no one person can, can do everything that's needed. So there's collaboration and teamwork required. And then ultimately, we need to publish our results that include the models in a reproducible way for sort of trust part of the open data movement. And I think this applies to many National Science Foundation disciplines. Ultimately, we have to be about helping solve the grand challenges of hydrology or whatever the field it we're working in. And a grand challenge for hydrology is data forecasting that quantifies the effects and consequences of land surface change on hydrologic processes and conditions. And I think that we can help with this by enabling access to and organizing the data for integrated analysis and modeling. The way the data's organized, and the data models that we use can either enhance. Or if it's disorganized, inhibit analysis that can be done. So working on the [inaudible] the data models to enable access to data is an important part of, of hydrologic modeling and hydrologic forecasting. So that we can better forecast looking at the picture in the bottom left. And plan for and mitigate the effects of floods. So that we're not necessarily picking people off roofs in an emergency situation. Or we can ensure sufficient water resources, looking at the picture at the bottom right in times of, in times of shortage when our reservoirs may get really, really low. So I'm now going to show a few examples from HydroShare before I get into describing exactly how it works. So this is a screenshot of a, of a paper from the Journal of Hydrology by Jeff [inaudible] coauthors. Looking at flood severity and modeling flood severity using crowdsourced environmental data. And I wanted to just highlight this to point out that on the right we've got an analysis where he's looked at flood reports. And then the predicting the number of flood reports based on the data-driven method that he's evaluating. But in the citations to the paper at the end, there are [inaudible] identify citations to the input data for this analysis that are saved in HydroShare. The output results that are also saved in HydroShare as well as the, the model scripts. So this is an example of where one can move beyond only the paper that's reporting the findings to the, the data that's needed to reproduce the findings. And there's a linking from the paper to the data. And if you go into the HydroShare resources here, you'll also see in the references part of that, there's links from, from the data back, back to the paper. So it's an example of linking research product, products and the, and the data. And I think that's an important element of closing the loop for, for research reproducibility. Another example, and this was with the help of National Science Foundation record grant is collections that are being assembled from the 2017 hurricanes that impacted the sort of southeastern coast and southern coasts of, of the US. Harvey, Irma, and Maria. So while they're still effectively works in progress. But you can search in HydroShare for those keywords and start getting some information. And we're hoping that the collaborative aspects of HydroShare are going to allow the community to benefit from these collections as they study these hurricanes. And figure out how to better design infrastructure and better work with systems to, to, to sort of mitigate a plan for presenting the catastrophic effects of, of those sort of events. So HydroShare supports a number of research communities. And this is one example where there's the land lab earth surface modeling tool kit that's part of the community surface dynamics modeling system. And that's [inaudible] infrastructure in its own right that's actually supported by the SR2 program. And what's happened with land lab and HydroShare is that the, on our Jupiter [inaudible] the Landlab software has been stored. A group in HydroShare has ended up getting about 330 Landlab users. And this is really helpful for, for people who want to use Landlab but perhaps don't have the ability or are being challenged the, the process of installing and getting all the, the libraries just right. So it helps you to get up and running quickly without them having to, to deploy software. So it's ideal for classroom and workshop use. And then it's also ideal for collaboration because when you've got data in HydroShare that comes out of a Landlab analysis, it's really easy to share with, with other colleagues. So it's been nice to work with effectively separate effort in, in advancing a surface, a surface modeling. One of the results from that is illustrated in this paper here. So [inaudible] got a paper that's published in this case Earth Surface Dynamics. And it's looking at landslide sources, debris flows, and the like. And I'm not going to sort of pretend to, to try and explain the figure on the right that is depicting the, the research that was done in terms of trying to predict landslides. But just points out that the, the data is again published in HydroShare together with the model code and the user's manual, all in the form of Jupyter notebooks. So I'll be talking in, in a bit about how we're using the, the Jupyter platform to enable computation and analysis from data in HydroShare. So HydroShare first went online in June 2015, and the graph on the top left shows the growth in the number of users. So at some level this is a, a measure of the broader impact. We also keep track of the users that come back. So we've got a graph there of the users who logged into HydroShare in the last 90 days. And also the new users each, each 90 days. And one of the things we're doing is effectively tracking, tracking use so we can see what use we have and who is using the system. So the pie chart on the right shows the partitioning between graduate students, faculty, commercial professionals, university professionals, post-docs and so on. The primary audience, at least in the way we are developing the system is the US [inaudible] research community. But this is a system that's on the web. It's on the internet. So anybody with internet access can, can use it. And we do have users from, users from other countries. And it's open, it's open for anybody. And we hope it becomes a community resource for sharing integration of data to address problems around the world. This slide also at the bottom left shows the number of resources. So we actually refer to the content, and I should have made that point a couple of slides back. The content in HydroShare where we've got different types, types of datas are referred to as resources. Because that includes both data and models. And we think of those resources as social objects. Things that people will collaborate on. So in this, coming back to the slide on the audience and the user base. We've got about 5500 resources. Only about 1500 of that looks like just less than a third of them are public. And that's because a lot of times people put data in the system and they may be using it just to collaborate with, with their colleagues while they're in the process of doing their work. Adding metadata to it, and then make it public later on. We believe that that's an important use case. Sometimes people say, "Well, why don't you require that the data always be, be public." But that actually would then mean people would only put data in the system at the end of the project. And we think it's much better for people to put data into the system as you're doing, as you're doing your work. So that you can add the metadata when it's, when it's fresh. And then, and then you end up with better metadata and then you make the information public when it's, when it's appropriate. So this slide lists some of the cyber infrastructure challenges involved in working with what is really a field of research data management. The picture on the left just really sort of depicts the fact that there's a computer and we're working with, with cyber infrastructure. But there's a, we all know about large data sets, the data deluge, [inaudible] of data. Challenges associated with inadequate metadata. So we're trying to have easy ways to address that. The way the data's organized. Model input data preparation. Reproducibility of computations as well as how well our models reproduce the physical processes they are intended to reproduce. Challenges associated with software installation. Libraries, platform dependencies. Sometimes licensing if you, if somebody publishes their results and it depends on proprietary software, then that may not be easily reproducibility. And you end up just having to sort of trust what they've done. So those are some of the challenges. And then computational resources as we deal with larger and larger data. So getting into data management and the requirements for data to be open and accessible is becoming increasingly important. Within the National Science Foundation and elsewhere. And this is a statement that came out of the, our Utah Project which was [inaudible] project in the state of Utah. [Inaudible] was the person who was picked to be leading the [inaudible] infrastructure part of that project. And he's also part of the HydroShare development team. And the data management states that all of the primary data sets will be collected. Collected, will be made freely and publicly available. How do we accomplish that? So an ideal investigator data workflow might be that you create a digital instance of a data set or model. Then you quickly share it with colleagues. That's priority at first. This is where the private data sharing comes in and where the collaboration comes in. Then there's value-added through the collaboration as annotation is, is added. And there's iteration. Comes described with metadata that's described all the work that's gone on to get the data set into the form where it's ready for publication. And eventually then you share it and perhaps formally publish it. It might get assigned a digital object identified and made mutable. One of the things that's motivating us is that it's not as easy as it should be. And HydroShare's intended to help with this process. And we're hoping that some of the solutions that we've been advancing in the field of hydrology will also translate to, to other disciplines with perhaps appropriate changes for data formats and nuances for the different ways different disciplines work. So the picture on the right shows the HydroShare website. And then the smaller box on the right shows landing page for a resource in HydroShare. As I pointed out earlier, HydroShare is a web-based hydrologic information system operated by the Consortium of Universities for the Advancement of Hydrologic Science, Incorporated, CUAHSI. It's, it's important to have a community that's helping drive the development and effectively provide the, the, the, the community of interplay between community and cyber infrastructure is key to make sure that what we do is actually meeting needs. And, and it gets used. HydroShare gives you a way to share data sets and models and other research projects. It's got the capability for collaboration. It has links to computational resources and it provides permanent publication of data and models with citable digital [inaudible] identifiers. If you look really closely on the bottom-right of this picture, you'll see that there's a DOI associated with this, with this resource. So all in all, it's a system to advance the field by enabling more, the community to more easily and freely share data projects resulting from the research. Not just a scientific publication, but also the data and models used to create the publication. And that fits into the increasingly popular paradigm that ADU is advancing. That came out of sort of the [inaudible] principles for data to be findable, accessible, interpretable, and reusable. That's a really a set of principles that the community's being encouraged to sign onto. So ultimately, HydroShare is a platform for sharing hydrologic resources and collaborating. You can think of it first as file storage which provides DropBox-ish type of functionality. And the picture on the right sort of depicts putting files in, in a DropBox to share with others on a computer. But adding value in terms of metadata descriptions. Adding value in terms of data, data access application programming interface. Web apps that can work on the data. Social functions for collaboration. And then [inaudible] to publication with digital object identifiers. So that's some of the value-added functionality we're trying to integrate into the system. I pointed out earlier. Many of the, the, the formats used, time series, geographics maps to geographic features. Multidimensional space-time data. Each depicted with one of the sort of picture on the, on the right. But also importantly, beyond data, moving into model programs and model instances. And we draw a distinction between the model program which would be the computational program that actually does the calculations. And the model instance which is the input data and the output data from application of that program at a specific location. To try and represent the physical processes there. So we needed to come up with a way to hold all this information in a, in a general way. In sort of a reproducible way. In a structured way. Ultimately in a machine-readable way. So the data model that's been designed to do this is depicted on this slide here. Is based on the open archives initiative object reuse and exchange standard. And the citation at the bottom is for the paper that describes it. And it really provides a way to hold the information for our resources, our data in the different formats in the, in an organized and machine-readable way. Where each resource may describe a number of aggregations that are effectively groupings of information held within a resource. And then each aggregation can hold resource-level data. Might be in the bottom-right here we've got a [inaudible] file holding observation data model two time series data format. It may have a title. For example, water temperature, and a little bit of a [inaudible] so there's a structured way of representing the information that's, that keeps things, keeps things organized. While this is, if you show a picture like this to users, they, they do get intimidated and we've really focused on trying to make the interfaces simple so users don't really have to know that this is all going on behind the scenes. But I wanted to point this out to show that there is a rigorous data model behind this that could be used by other fields. So this slide here shows the My Resources page which is a page in HydroShare that you use to add new information. So click on the create a new resource button. And then you'd add content to HydroShare. And once it's edited as the resource, then you would be able to share it with your colleagues or permanently publish it. This shows the resource landing page. Again, what I've got depicted here is the screenshot that shows the, some of the basic metadata that's, that's part of the open archives initiative object research and exchange standard. But presented here in hopefully a pleasing way for, for users, for the authors, the owners, the type of the resource. Things like the date and time which was create an abstract, how to cite information, licensing information. It's important when information is published on, on the web to be clear about users can, can do with it. And we've tried to build in the capability for users to specify which license to use. We are encouraging the use of Creative Commons but users can actually put whatever license they want. And they can also require that when other people download the data, they acknowledge that they have read the license and agree to abide by it. For each resource, you can manage where other users can edit it or just view it. You can, there's commenting and rating capability. There's a lot of attention to unique identifiers, and we're using globally unique identifiers here for, for all of our resources. And describing with metadata, you can organize resources into collections. You can permanently publish it. You can create different version. Of course once something's permanently published, you don't want to change it. But if there's an update, you might want to create a new version. And then you can open it with compatible web app. And I'll be talking about that in a bit. So I've mentioned that the objects that we work on in HydroShare are referred to as resources. And resource can hold multiple aggregations. And here I've got on, depicted in the picture on the right these resources that might hold multidimensional data. It may hold geographic features. It may hold geographic grid information. Each resources is effectively managed as one discoverable object with one set of access and the owner--access-controlled information. One [inaudible] resource-level metadata, things like the abstract title, the keywords. You can create collections of resources. And that's what we have done for the hurricanes Harvey, Irma, and Maria. And collections and their members may be discovered separately. You can also form informal collections through the use of keyword tags. So for example, I'm expecting there's only one resource with this particular keyword, OAC Webinar. But if you say are going to a conference and you want everybody just to put all the, all the data or perhaps the presentations into HydroShare and agree to use the common keyword, then it becomes a way to discover those things. And as long as there isn't a clash in terms of keywords that people choose. It becomes an informal way of building connections. We're trying to keep the metadata entry as simple and easy as possible. Recognizing that metadata's a thing that often researchers see to be painful. So wherever possible, we will read the metadata from the data files themselves. And depending on the data, the format of the data files, that can be done to a greater or lesser degree. If we're getting data from multidimensional [inaudible] format file that's followed the [inaudible] conventions. Which is a [inaudible] conventions that is promoted by the [inaudible] science community. Then the metadata can be harvested and automatically entered into HydroShare's metadata records. And the user doesn't have to do very much. Similarly with Geotech information which is a format for gridded data. There's the, the GIS information can be, can be read. The G--the geographic information system information. So that has been as much as I wanted to say about how information can be saved and shared in HydroShare. What I'm not going to talk about is actually working with data in HydroShare. Because we learn from our users that there's a need to be, need for more than simple data sharing. Users want to work with their data that's actually provides an incentive for people to put their data in the system and, and therefore sort of do the right thing in terms of eventually publishing, publishing their data. So this has led to the concept of web applications or web apps, tools that can act on data right in HydroShare. So if you click on the apps tab on sort of top menu of HydroShare, you'll see listed some of the apps in the, in the apps library. And you'll see that there's a geographic information system app. The Jupyter [inaudible] app that opens Jupyter notebooks that I'll be talking about. An app for viewing forecasts from the national water model and then the [inaudible] explorer app is one for viewing some NASA, NASA data. So to understand how apps are all sort of configured, I just wanted to explain a little bit about the conceptual architecture that HydroShare uses. And this picture in the middle here really shows three parts. A few other things thrown in. Think of the users at the top. That's these little orange depictions of, of people. The screenshots I've been showing come from the website that's implemented using the Django framework that Python framework for building, for building websites. And the data's held in iRODS which is a distributed, fairly advanced file storage system that allows you to [inaudible] around different locations on the internet for where the, the data's stored. So the, the Django website supports the capability for discovering information, for interacting with the metadata. It's the resource exploration functionality. It's where you can organize and annotate your content, manage access, and so on. iROD is similar to the hard drive on a computer. And it supports distributive storage. But then on the sort of top-right of this diagram. I've got the depiction of apps. And those allow you to do program actions or analysis and modeling on the resources. And it's web software to operate on the content. And it's all connected together through application programming interfaces. So we've put a lot of energy into have the API have the capability to access content from, from HydroShare resources. And one of the ways we've tried to do this. Anybody can set up a server or an app or a platform of software servers. To operate on HydroShare resources. So we do not want to limit it to the apps that we can develop. And we see that the ability to have other people, third parties develop apps is a major mode for, for extensibility. And there's an examples of a few of the apps. Some of them are listed here. SwapShare is one that's actually hosted at [inaudible]. I'll talk about the Jupyter hub is actually, we're hosting it. We can use the thread server from Unidata to access multidimensional information and so on. So to configure an app to, to be accessed, to access information from HydroShare, this is, this slide depicts where that linkage mechanism, where there is a resource type in HydroShare that's referred to as a web app resource. And it holds the metadata for launching any third-party website. So if you create a web app resource, you can then go in and specify what the URL is to launch that app. It could be for example apps.Hydroshare.org if it's one of the ones that the HydroShare team has written. Or it could be mygeohub.org which would be the one for SwapShare hosted at Purdue. And then you can specify things like the resource identifier. The user name that need to get passed to that app. The, the sort of information that the app needs to be able to access the data from HydroShare. Using the API. So to [inaudible] general construct that allows anybody to establish this capability. So next this is one of the ways that we are intending for HydroShare to become part of a broader ecosystem, taking advantage in levering, leveraging capability of, of other software that people develop. So I'm now going to go through an example of some of the apps and what they can do. We've been putting a lot of energy into, into Jupyter hub because that provides a fairly simple to use and documented, reproducible way for, for a workflow to be sequenced. So if, if I'm here looking at a HydroShare resource, and that's what's depicted on sort of the main part of the screen with the hydrologic terrain analysis Jupyter notebook. That's a file that's stored in HydroShare. It's got authors, owners. There's an abstract that describes it. And in its content section, there's just a single file that by the extension IPY. And B, we know that it's a Python notebook which is the format that Jupyter notebooks can use. So if you click to open this resource from HydroShare with the JupyterHub app, you end up switching now to the, the separate system that is hosting JupyterHub. And you can open that notebook that's been provide by a HydroShare user. In that case, in this case the user was me. But it could have been any user. And provide the ability for, for another user to step through the analysis that's documented in that Jupyter notebook. So you can write and execute code in a Jupyter notebook, acting on the contents of the HydroShare resources. Save the results back to the HydroShare repository. So for example, I've got depicted in the bottom here. The added content to an existing resource. That looked at files that were the result of whatever the computations were that this Jupyter notebook, this Jupyter notebook [inaudible]. So this is one mode for doing an analysis that is supported in HydroShare. Now I recognize that this is a complicated slide. One of the important ideas that we're trying to advance in hydrologic modeling is need to systematically evaluate multiple working hypotheses about the model representations of physical processes. So SUMA, or the Structure for Unified Modeling Alternatives, is an approach that's being advanced by Martin Clark and these many authors published in Water Resources Research. To effectively systematically evaluate, as depicted on the right here. Considerations like what good representations should we use for our hydrologic model? Should we break it into a grid? Or for example into watershed-based polygons or single [inaudible] that represents the whole watershed. How should the vertical processes be organized? Single soil column or soil column on top of an aquifer? Or two separate soil columns on top of an aquifer or some sort of a hill slope depiction? So SUMA effectively providing a way to systematically evaluate which of those different representations compare better in terms of model simulations against, against observations. It's a system right now that's somewhat intimidating and, and hard to use. So we've been working on building it into the Jupyter notebook platform to effectively make it more accessible and easier. So this is an example where there's a reproduction of some of the results from one of the papers that initially published SUMA results. So there was an examination of different representations of way to calculate the model resistance in evaporation. And the, the Jupyter notebook environment allows you to specify them in a simple set of text or text functions that you enter into the Jupyter notebook. And then you can effectively run the analysis to, to test each of those different simulations. And you can produce the, the graph that, that's similar to a graph that's in a, in a plotted paper. We're also working in this next slide to produce enhanced geospatial data in Linux capability through effectively Jupyter notebooks that are attached to higher performance computing systems. So this is some of the work with [inaudible] and his colleagues at University of Illinois. I've got noted in here that this is in progress. Because we don't, this is not actually deployed on HydroShare right now. But this is an proof of concept illustration that we will be deploying in the future. Then there are web apps that enable perhaps more simple visualization of things like time series. So there's a quasi-data series view that enables that. And then I'm going to go quickly through the next set of slides. They are apps that facilitate and ease access to data sets that are of community value. That may not be that easily accessible otherwise. So for example is the app to provide an interface to national water model app. So this is a water model being run by the National Water Center for operational hydrologic forecasting. And there's a need for researchers to have, have access to this, this information. There's a, an app for viewing [inaudible] smoke cover data. Also again makes it easier for people to use. And there's an app for accessing the gravity recovery experimental or GRACE information which tells you something about groundwater which can for example depict the droughts in California. As the, as you get less and less water, the gravity field is impacted. I just wanted to talk a little bit about some of the infrastructure that's involved. So HydroShare is hosted at [inaudible] Computing Institute in North Carolina where [inaudible] and the team there have really worked hard on setting up a system that can support the sort of continuous integration. The contributions from different developers. The automated testing. And we've really had to work to effectively train developers that range from graduate students to software professionals on, on getting, getting code into and tested by the system. But we've relied quite a lot on Jenkins and the automated testing system. And then the storage system that provides backups and replications [inaudible]. So here's a little bit of software engineering process. Sort of collaborative open development process used by a team that ranges from professional programmers to graduate students and part of I think one of the success but was also challenging. Was getting everybody involved with using GitHub for issues. Branching [inaudible] requests, code review testing the emerging and applying. We've also worked to try and make the user experience good. And here we have we subjected HydroShare to a user experience redesign from the [inaudible] Capstone Project. So we had a group of five students. And there they're shown all holding, holding big balls. To really look at the, the usability of HydroShare from the sort, sort of from, from a fresh perspective and from a user-experience perspective. Talking to users, going through the process of code design, coming up with the personas are our sort of, characterize the typical HydroShare users to help with the design. And they produced a nice design report. The full report of which is in this HydroShare resource. And provided, for example, recommendations on adding a dashboard to HydroShare. Adding capability for a, a search bar that searches all of HydroShare. And we're going to be over the next few years of this project, really working to bring in some of the capability to enhance the user, the user experience. So this slide tries to show just a summary of what we're working with. We're working towards with HydroShare. As a web-based collaboration environment to enable advances in hydrologic understanding through collaborative data and shar--data sharing and analysis. Really integration of information from multiple sources. And there's two parts to it. There's the repository part of it, which provides sharing and publication of data. The social discovery and added value, sharing of models. And that's all hitting the points of collaboration reproducibility, giving people credit for the data they collect. That's another important part. And transparency. And then there's the computational part. The modeling and analysis at this depth can be facilitated through the system. For things like standard ways of model input preparation, model execution, perhaps best of practice tools for visualization and analysis. And this is where there's connectivity and interoperability with server and cloud-based computational systems that can take advantage of interoperability in using other systems. So this is my last slide. I wouldn't have been able to do this without the help of all of those people underneath umbrellas in the rain. [Inaudible] meeting and it's operated by CUAHSI as a data facility to try and effectively sustain it for the hydrology community. And if you want to learn more, the website is, is there. And there's also a set of help pages and [inaudible] pages where you can get publications and things like that. So thank you.

>> David, thank you so much. Let, let, let me begin the Q&A session by asking you sort of asking you a broad question. So could you talk about some of the key challenges that you and your team overcame in making HydroShare a shared cyber infrastructure? Basically that could serve as takeaways for others seeking to build this kind of large scale [inaudible]?

>> Okay. So I think the probably there's quite a few of them. One, one of, one is the disciplinary interface between computer science and software professionals. As well as, and well then the, between that and, and hydrology researchers that's required us to really work on communication. Work on things like the use cases and learn the jargon of, of each field. And be patient in, in the communication. It's required us to be careful about sort of distilling the overall variety of wants into sort of prioritized and actionable capabilities. And identifying and setting priorities that sort of address use cases. We want to be able to balance adopting existing systems. Adapting them when necessary with developing new functionality in a way that assembles a system that, that meets needs. So we've tried to use things like Django as a web framework. That sort of hopefully at the right level that our team can, can develop with. It's got the flexibility. It's not quite as rigid as perhaps a high-level framework would be. But it does reuse existing capability. So those, those are challenges that we I guess continue to, to grapple with. And then I guess user experience, making it simple and natural for users. Which was one of the reasons for getting the [inaudible] team involved.

>> Thank you. So there's a question on the chat box. Is there a quality screening process in the overall data publication workflow?

>> So that's a good question that we always, often get asked. Right now there is not. In the sense that anybody can go and load data. And, and get it published with a, with a digital object identifier. The system will not let you do it if you haven't filled in all of the metadata. But it doesn't actually read the filling in of the metadata. We've identified this as a, as a limitation. So we have in the works, and it's almost complete really, ready to go out probably not the next release, but the release after that in about a month. Where we will have, [inaudible] performing a completeness check on the metadata. So that will provide some level of quality checking. Really just to, to demonstrate that, that the abstract actually makes sense. The files that are there seem to be complete and described by the metadata that they're, that we've avoided. We've avoided trying to build a system that gets into review of resources for, for quality. Because we would see that as inhibition to people using it. And that might actually then make, make people perhaps put their data in other systems that we would--. I mean, for example there's ThinkShare just allows anybody to put data in there. And I don't think it provides quite as [inaudible] organized metadata as HydroShare does. But it will still give you a

digital object identifier. And we would rather encourage people to use, use HydroShare for hydrologic data of, so they can take advantage of the capability that HydroShare is providing. So we don't want to provide too, too much of a barrier to entry.

>> Thank you. Let's go to a general sustainability question. Where do you get funding for operations? And training for operations and training new users when you know most of the funding is for research? In general, what do you see as a viable route for sustainability for HydroShare?

>> Okay, well we don't have enough money, and probably everybody would say that. The funding right now is coming from the National Science Foundation through CUAHSI, so CUAHSI is supporting the, the operation of, of HydroShare. We're trying to make the system simple so that not a lot of training is required. We've partnered with other projects wherever possible. And so for example, the, the IUtah project that I showed the example of the data management plan. That's provided training for the IUtah users to put their data into, into HydroShare. And some of the training tools and material that they developed benefits the, benefits the rest of the project as well. In terms of, in terms of long-term sustainability, we are working on a way to minimize the costs by having data in third-party storage. So that's one of the, the major objectives of our current development. And that would include things like the large storage systems such as Box that many universities have bought into as well as digital libraries. I think we ultimately are going to have to get to where we find acceptable ways to have users pay for some of their services. But that's, there is a challenge where a lot of users expect everything, expect everything for free. And it's not a problem that's fully solved yet.

>> Thank you. And really, thank you for your presentation. We are at the end, we are at the top of the hour. So thank you so much, David. Any words from you, Amy?

>> David, thank you so much for joining us today. And of course thanks to all of our participants. For staying online and for listening so attentively.

>> Well, thank you for having me. It's been--.