

# Increasing the Accessibility of Computing Cyberinfrastructure in Hydrology by Streamlining the Integration of HydroShare with CUAHSI JupyterHub

Kyle Combes, Emily Lepert, Vicky McDermott, Charlie Weiss

Olin College of Engineering SCOPE senior capstone partnership with the Consortium  
of Universities for the Advancement of Hydrologic Science, Inc.

May 2020



# Executive Summary

CUAHSI Compute is a free collection of web-based computing platforms that are actively developed and maintained by the Consortium of Universities for the Advancement of Hydrologic Science, Inc (CUAHSI) with the goal of making cloud-based education and water science research more accessible to members of the hydrology community. It offers a means of running code written in various software languages right in a web browser without any laborious or frustrating setup process. Additionally, its integration with HydroShare, a water data and model repository, allows users to easily reproduce work done by others and smoothly share their own work with others. The oldest and most widely used compute platform of this collection is called CUAHSI JupyterHub.

While CUAHSI JupyterHub offers great promise, its evolution thus far has centered largely on technical functionality. Ease of use, including the user interface and user experience, has yet to be fully refined, particularly with regard to the integration with HydroShare. This has left the system in a state where it is accessible to its creators but much less so to new users. CUAHSI recognizes this shortcoming and has partnered with an Olin College of Engineering SCOPE team in order to address it. The Olin College of Engineering SCOPE program is a senior capstone program which seeks to give students experience working on “real-world” problems. Groups of four to five undergraduate engineering seniors are paired with a sponsoring company or organization for one academic year to work on a project with implications beyond Olin. We, the 2019-2020 CUAHSI SCOPE team, have spent the past eight months working with CUAHSI in order to use our experience in user design and software development to improve the user experience of the CUAHSI JupyterHub platform.

To identify the needs of potential users and the current gaps in CUAHSI JupyterHub, we interviewed hydrologists and conducted co-design activities both virtually and in-person. We then used our insights to develop a web app, called CUAHSI JupyterHub Sync, that better integrates the CUAHSI JupyterHub platform with the HydroShare repository. The app allows users to easily work with a HydroShare resource’s data with the tools in CUAHSI JupyterHub and to intuitively synchronize data between CUAHSI JupyterHub and HydroShare, making the integration of CUAHSI JupyterHub with HydroShare more seamless. It also drastically reduces the number of steps needed to synchronize data between the two from six or more steps down to only a single step. An alpha version of this app is in the process of being deployed by CUAHSI for preliminary testing by members of the hydrology community.

In summary, our user research has unearthed novel insights into how hydrologists interact with existing cyberinfrastructure and what needs are still going unmet. We used these insights to develop a working prototype of CUAHSI JupyterHub Sync, which provides an improved data management interface for the JupyterHub component of CUAHSI JupyterHub. We have taken the first steps toward implementing a more seamless web-based computing experience and believe we have laid the groundwork for CUAHSI to take its offerings to the next level.

# Acknowledgements

We acknowledge funding from the U.S. National Science Foundation for OAC-1664061 (Bandaragoda, Tarboton, Black) and CUAHSI EAR-1338606 (Pollak, Castronova). These development grants support the Consortium of Universities for the Advancement of Hydrologic Science, Inc (CUAHSI), University of Washington, Utah State University, and Olin College of Engineering. Other contributors to the development of the HydroShare platform (<http://www.hydroshare.org>) include Brigham Young University, Tufts University, University of Virginia, University of Washington, National Center for Atmospheric Research (NCAR), Purdue University, and University of Texas at Austin. We acknowledge early contributions from HydroShare developers supported by ACI 1148453 (2012 - 2018). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We recommend users of this tool use this form to provide them information on acknowledging CUAHSI resources.

Partner grant details include:

USU HydroShare 1 2012-2017, ACI 1148453

RENCI HydroShare 1 2012-2017, ACI 1148090

USU HydroShare 2 2017-2021, OAC-1664061

UNC Chapel Hill, RENC I HydroShare 2 2017-2021, OAC-1664018

University of Illinois HydroShare 2 2017-2021, OAC-1664119

CUAHSI Cooperative agreement 2014-2020, EAR 1338606

USU Hurricane Harvey and Irma RAPID 2017-2018, OAC-1761763

We would also like to acknowledge the support provided by Christina Bandaragoda and Jon Pollak throughout the project. They guided us as we learned more about the world of hydrology, provided us with contacts, liaised with CUAHSI and others, and supported us in many other ways. We would also like to thank Tony Castranova and Scott Black for assisting with the integration of our new system with the existing CUAHSI infrastructure. We wish to acknowledge David Tarboton for his feedback and input throughout the project. At Olin College of Engineering, we would like to thank Lynn Stein, who served as an advisor on this project. We'd also like to thank Alison Wood, Amon Millner, and Tim Ferguson Sauder, who all gave invaluable input and advice at various stages along our journey.

# Table of Contents

<b>Executive Summary</b>	<b>2</b>
<b>Acknowledgements</b>	<b>3</b>
<b>1. Introduction</b>	<b>6</b>
<b>2. The Vision for CUAHSI JupyterHub</b>	<b>7</b>
<b>3. Current State of CUAHSI JupyterHub</b>	<b>7</b>
<b>4. Identifying the Next Steps</b>	<b>9</b>
4.1 Conducting Interviews	9
4.2 Key Insights	10
4.2.1 Hydrologists' Workflows	10
4.2.2 Potential for Use in Education	12
<b>5. Building a More Integrated Experience</b>	<b>12</b>
5.1 User Interface and Experience Design Elements	13
5.2 Web Technology Stack	14
5.3 HydroShare Jupyter Sync Web App	15
<b>6. Next Steps</b>	<b>18</b>
6.1 Remaining Issues	18
6.2 Educational Resources	18
6.3 Collaborating with HydroLearn	19
6.4 Sync Between Resources	19
6.6 Access CUAHSI JupyterHub Without Using HydroShare	19
<b>7. Conclusion</b>	<b>21</b>
<b>References</b>	<b>22</b>
<b>Appendices</b>	<b>22</b>
A. Initial User Research	23
B. User Research at the 2019 AGU Fall Meeting	24
B.1 Moving a Resource Activity	24
B.2 Workflow and Mood Card Activity	24

B.3 Web App Walkthrough	25
B.4 Survey	26
B.5 General Interviews and Discussions	27
C. Directory Exists Prompt	27

# 1. Introduction

The Consortium of Universities for the Advancement of Hydrologic Science, Inc (CUAHSI) is a nonprofit organization that builds and maintains computing tools for hydrologists, or scientists whose work centers on the study of water. While technically a consortium of researchers from across the globe, CUAHSI has a small full-time staff based in Cambridge, MA.

Today, CUAHSI is heavily focused on equipping hydrologists with the tools they need to succeed while also improving the reproducibility of hydrologic science as a whole. CUAHSI's HydroShare platform serves as a repository for sharing and archiving hydrologic data and models. It allows anyone to upload their code, water data, and any other accompanying files into a "resource," which is assigned a digital object identifier (DOI) to allow easy referencing in publications and the like. People can also add others as collaborators on a resource, thus helping to facilitate collaboration among hydrologists. In addition to HydroShare, CUAHSI is creating a web-based suite of tools for computing called CUAHSI Compute. The oldest and most widely used of those tools is CUAHSI JupyterHub, which allows hydrologists to easily run Python and R code right in their web browser. By providing easy access to a web-based computing environment with virtually no setup hassle, CUAHSI JupyterHub can significantly reduce the barrier to entry for computing. Being cloud-based, it also offers the potential for increased computing capacity and faster code run times. These two tools, HydroShare as a data repository and CUAHSI JupyterHub as a computing tool, equip hydrologists with a solid computing toolkit for conducting their work.

HydroShare and CUAHSI JupyterHub have great promise, but they have yet to attain their full potential. CUAHSI developers are spread thinly across many projects, and thus they have not always had the time and energy needed to work on aspects of the projects beyond their core functionality. CUAHSI's vision is for HydroShare and CUAHSI JupyterHub to operate seamlessly together, one providing data storage and the other a workspace for computing. So far the integration of these two platforms into one cohesive usage experience has been limited, but CUAHSI recognizes the value such refinement can bring to their system.

This is where we, the Olin SCOPE team, come in. The Olin College of Engineering SCOPE program is a senior capstone program which seeks to give students experience working on "real-world" problems by pairing with a company or organization for one academic year. As Olin students, we are trained in user-centered research and design as well as software development. This positions us well to assist CUAHSI in taking the JupyterHub component of CUAHSI JupyterHub to the next level by conducting user research, proposing design changes, and prototyping these changes ourselves.

Over the past academic year, we interviewed hydrologists, drafted user interfaces, solicited feedback on the drafts from potential users, and finally developed a working prototype web app to better integrate HydroShare with CUAHSI JupyterHub. We have only taken the first steps toward implementing a more seamless web-based computing experience, but we hope we have given CUAHSI what it needs to take its offerings to the next level. We see the promise in HydroShare and CUAHSI JupyterHub, and we hope our work and recommendations help CUAHSI realize that potential.

## 2. The Vision for CUAHSI JupyterHub

One of the primary goals of CUAHSI is to help advance the field of hydrology by making it more open, collaborative, and supported. CUAHSI hopes to increase the verifiability and reproducibility of hydrologic research by making data more accessible, and to increase trust by making the means of discovery more transparent. CUAHSI also aims to foster collaboration among hydrologists by providing an easy means of sharing data and collaborating on models. Additionally, CUAHSI hopes to empower hydrologists to conduct their research more quickly and effectively using powerful, easy-to-use, and hassle-free compute tools.

CUAHSI recognizes that current computing systems often present significant hurdles and limitations, especially for those who are not experienced computer programmers. Just getting started with Python or R requires installing and configuring software, dependencies, and environments. Even if a hydrologist is able to configure Python or R on their computer, they may find their ability to conduct research limited by their computer's hardware. Running complex models often demands more compute power than a consumer laptop or desktop can handle. Thus, a way to run models on a powerful, setup-free computing system may help numerous hydrologists perform more complex work.

CUAHSI JupyterHub has the potential to be this tool, as it offers a complete, easy-to-use, and powerful compute system to hydrologists. CUAHSI JupyterHub allows users to run Python and R code right in their web browser with no setup. This eliminates the setup hurdle that many first-time users of Python or R face, thus reducing the barrier to entry for new users.

CUAHSI JupyterHub can also integrate seamlessly with HydroShare, which alleviates the tension in data management by allowing data and models to easily be shared with others. With a well-designed user interface, a HydroShare resource (which is simply a collection of files and folders) could be opened in CUAHSI JupyterHub with the click of a button. This would make it easy for those with little to no programming experience to re-run another hydrologist's model, inspect it, and verify the output for themselves. Together, HydroShare and CUAHSI JupyterHub have the potential to streamline the process of managing and sharing data and models, thus enabling easy collaboration between hydrologists.

## 3. Current State of CUAHSI JupyterHub

A solid foundation for CUAHSI JupyterHub is currently in place, offering a robust and reliable way for users to run Python and R code in their web browser. Users can open a HydroShare resource in CUAHSI JupyterHub by simply clicking a button labeled "Open with CUAHSI JupyterHub" on that resource's page in HydroShare. The files in the resource are then displayed, and the user can open and run Jupyter notebooks.

However, the format in which this is displayed does not make it clear to new users what they should do next. The files in the resource are simply listed, with no instructions on how to actually open a

notebook, or which notebook (if there are multiple) should be opened (Figure 1). This makes it difficult for users who are unfamiliar with JupyterHub to start using the system.



Figure 1: Currently, if a user clicks “Open With” and “CUAHSI JupyterHub” on a resource page in HydroShare, they are redirected to a page displaying the contents of the resource. While users familiar with Jupyter may know what to do next, new users are likely to struggle to determine the appropriate next step. This is especially true for resources with many files.

Additionally, if a user has made changes to a resource, there is no easy way for them to save that change to HydroShare. Two options currently exist (Figure 2), but neither is quick or straightforward. One option is for the user to download the data or notebook to their computer, open the page for the resource on the HydroShare website, and re-upload it there. This requires performing at least six clicks on three different pages, with a significant amount of time spent waiting for pages to load and scrolling up and down looking for buttons and files. That click and page count is the minimum number required and assumes the window for selecting the file to upload to HydroShare opens to the same folder on the user’s computer that the file was downloaded to from CUAHSI JupyterHub. This would likely not be the case during normal use. The user would probably need to navigate their filesystem in order to locate the file they want to upload. If this were the case, then the click count could easily be doubled. This higher count is probably more representative of the average user and use case, making this file upload route an extremely lengthy and tedious process. The other option for a user to save changes to HydroShare is for the user to write Python code that incorporates the `hs_restclient` Python module [1]. This module serves as a wrapper for the HydroShare API [2], and it allows users to initiate file transfers and manipulate the resource in other ways. However, this latter method in particular is not accessible to users who are not comfortable with Python. A faster and more accessible means of transferring data between CUAHSI JupyterHub and HydroShare is needed.

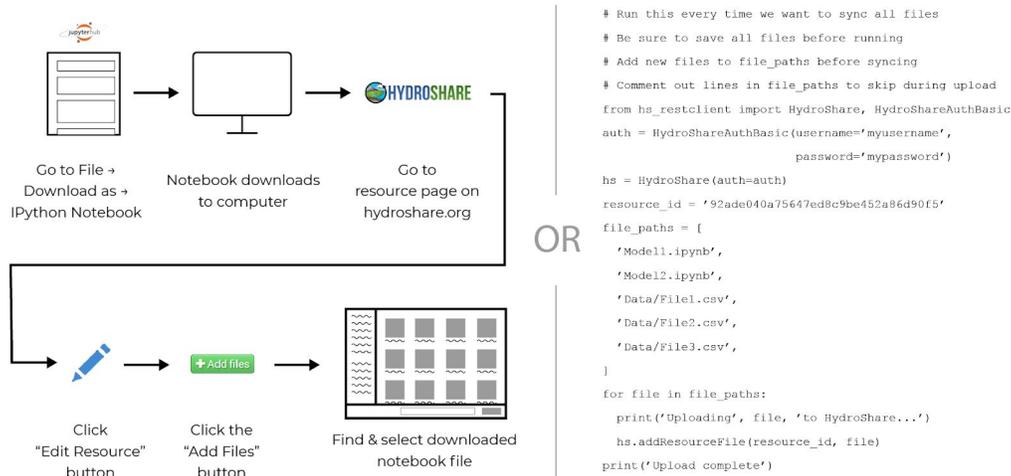


Figure 2: Two options currently exist for syncing changes in JupyterHub to HydroShare. The user can download the file to their computer, go to hydroshare.org, and then upload the file to their resource (left). The user can also use the `hs_restclient` Python module (right), but this requires referencing the `hs_restclient` documentation and writing Python code. This latter approach thus presents a hurdle for novice users.

The current workflow for re-opening a resource in CUAHSI JupyterHub is also a multi-step process with the potential to accidentally cause irreversible data loss. First the user must visit the resource page on HydroShare, and then they need to click the “Open with CUAHSI JupyterHub” button. At this point, if the user has opened the resource before, a popup appears telling the user that the “Directory Exists” and prompting them to either overwrite it or to go to the existing directory ([Appendix C](#)). While it may be obvious to the experienced user that the correct choice is to click “Go to Directory,” it is very easy for novice users to get confused and click “Overwrite.” If they were to do this, any changes they had made and not saved to HydroShare would be permanently lost. We did this multiple times ourselves when first learning to use HydroShare and CUAHSI JupyterHub. Ensuring that there is a way to re-open a resource without creating such confusion or loss of data is a minor change that would significantly improve the user experience of CUAHSI JupyterHub.

While a solid foundation for CUAHSI JupyterHub is in place, some targeted refinements could likely greatly improve the overall user experience. These refinements would substantially benefit novice users who are just beginning to leverage the power of computing in their work. Reducing the number of steps to synchronize resources should be a top priority, as should ensuring that the user does not accidentally delete their data. Identifying the best alterations to make to achieve these goals required consulting CUAHSI JupyterHub's potential user base.

## 4. Identifying the Next Steps

CUAHSI invited the 2020 Olin College of Engineering SCOPE team to refine the user experience of the JupyterHub component of CUAHSI JupyterHub. To be most effective, we began our work by conducting user research and prototyping possible improvements. The eight month-long partnership allowed us to leverage our user-centered design experience to identify the needs of potential CUAHSI JupyterHub users and determine how CUAHSI JupyterHub might be designed to best serve

those needs. Our software development skills also allowed us to prototype suggested improvements, providing CUAHSI with an alpha version of an improved CUAHSI JupyterHub experience by the end of the partnership.

## 4.1 Conducting Interviews

We conducted user research during two phases: at the outset of the academic year during the months of September and October and at the American Geophysical Union (AGU) Fall Meeting in San Francisco in December. We first cover the methods we employed to conduct research during both phases. We then discuss the results gathered during both phases in the following subsection.

We began our research by speaking with as many hydrologists as we could. We leveraged CUAHSI's connections as well as those of our Subject Matter Expert, Olin Professor Alison Wood, to talk to 11 hydrologists during the months of September and October. Their backgrounds ranged from current graduate students to recently retired long-time professionals. Students and faculty were recruited from Arizona State University, Penn State University, the University of Colorado, the University of Texas, the University of Virginia, and the Colorado School of Mines.

Table 1: Our first interview phase consisted of speaking with 11 hydrologists at various stages in their career. We sought information to better understand hydrologists' workflows and how coding fit into them.

Stage of Career	Hydrologists Interviewed
Graduate Student	4
Postdoctoral Researcher	2
University Professor	2
Industry Professional	2
Retired Professor	1

Based on the University of Washington research guidelines [3] and the advice of our liaisons, we determined these interviews did not qualify as research requiring Institutional Review Board (IRB) review. This was primarily because our interviews were not a systematic analysis designed to contribute to generalizable knowledge of the computing needs of hydrologists. We considered our work to be a single isolated case study. Had our work been part of a larger study of more than three cases on research platforms in general, then it would have been considered human subjects research and have required IRB approval.

Interviews were conducted primarily via Google Hangouts and lasted between 30 minutes and an hour. We began by asking a prepared set of questions, which included how they entered the field of hydrology, what their current professional position was, and how they used computation in their work (Appendix A.1). We initially sought to identify commonalities among workflows, but after a few interviews we began targeting hydrologists' experiences learning to code. The reasons for this shift and the insight gathered from these education-related questions are discussed in Section 6.2.

We then used the insights gained from our non-education-related questions (discussed in [Section 4.2](#)) to develop mockups for potential improvements to CUAHSI JupyterHub. These mockups served as tools for soliciting feedback on our ideas and designs during later interactions with potential users.

In December, we attended the American Geophysical Union Fall Meeting to conduct more interviews and to gather feedback on our mockups. We did this by conducting various prepared activities with attendees in order to learn more about hydrologists' needs, the type of vocabulary they use when thinking of computing, and their impressions of our mockups ([Appendix B.3](#)).

## 4.2 Key Insights

Our interviews of hydrologists provided us with valuable insight into hydrologists' workflows, the tools they use, and their feelings toward those tools. We also discovered great potential for CUAHSI JupyterHub in the field of education. This section discusses those insights at length.

### 4.2.1 Hydrologists' Workflows

Through talking to over twenty hydrologists about their current workflows, we discovered much variation in the current landscape. Some interviewees used Python, others used MATLAB, R, or Excel. We attempted to identify common needs among hydrologists so that we could target those when designing for CUAHSI JupyterHub. However, we were unable to identify any specific common workflows or practices among those we talked to. High-level similarities existed, since all interviewees needed to acquire data, prepare it, and load it into their processing program of choice, but the similarities largely ended there. Sources, formats, and intended uses of data varied widely, which meant we could not identify more specific commonalities among the workflows of those interviewed.

Since the field of hydrology is so broad, we also noticed that hydrologists' needs and practices diverged more and more as they progressed in their careers. While many may have begun at a similar point, as graduate students with limited to no programming experience, the multitude of projects they work on shortly leads to more differences than similarities among hydrologists. This makes it difficult to design software to target such a diverse group of potential users. We have captured this observation of divergence in a graphical manner in Figure 3.

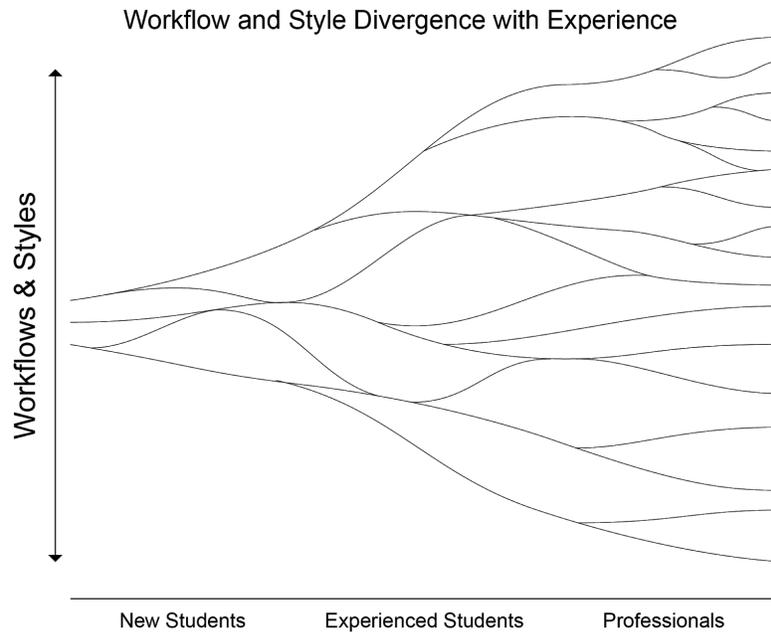


Figure 3: A conceptual model describing the challenge of developing shared tools in a diverse research community. We noticed that the further a hydrologist progresses in their career, the more likely they are to have a unique workflow and working style. This differentiation is due to the multitude of research environments, approaches to learning coding, and peer influencers that direct a hydrologist throughout their career. This divergence makes it increasingly difficult to build tools that are useful to a large fraction of experienced members of the hydrology community.

When asking hydrologists how happy they were with their current computing solution, we received a variety of responses. The general sentiment of more experienced hydrologists, who already used computing significantly and were comfortable doing so, was that they were happy with the tools they currently used. When we asked these hydrologists what CUAHSI JupyterHub would have to offer for them to use it, they were largely unable to name anything. One hydrologist expressed interest in a tool that would help convert his .py files into Jupyter notebooks for publication, but he was set on using his Spyder integrated development environment (IDE) with its debugger for actually writing and running his code. The rest simply remained silent or said they could not think of anything that would lead them to use CUAHSI JupyterHub as their primary programming environment.

Less experienced hydrologists, on the other hand, tended to face more difficulties with computing. We learned this more from speaking with educators than by speaking with inexperienced students themselves, as the educators were more readily available. As it stands, beginning hydrologists typically follow the advice of their more experienced colleagues when getting started with computing. Therefore, they tended to use whatever computing tools were already in use in their lab. Given this, we concluded that CUAHSI JupyterHub would first need to become popular with more experienced hydrologists before it would gain popularity among budding hydrologists. However, our conversations with experienced hydrologists suggested to us that this is unlikely to happen given their contentment with current tools, leaving inexperienced hydrologists to continue to struggle getting started with computing. Based on these insights, we believe that designing for inexperienced hydrologists who are just beginning their studies and making a concerted effort to advertise CUAHSI JupyterHub to these types of people could provide significant value.

## 4.2.2 Potential for Use in Education

We did discover that a number of educators are already using Jupyter notebooks in their curricula. Several professors were directing their students to use JupyterHub servers run by their institution. HydroLearn, an online platform offering educational content for hydrologists, was attempting to embed Jupyter notebooks directly in their web pages. We spoke with representatives of HydroLearn at the AGU Fall Meeting who said they were already using CUAHSI JupyterHub in their curricula. However, they expressed some frustration with the platform, particularly around the process the instructors and their students used to edit code in their notebooks. After editing files in CUAHSI JupyterHub, students are told to download their notebook to their computers, delete the old version from HydroShare, and re-upload the new version of the notebook to HydroShare for submission. This seemed unnecessarily long to them. They also wished they could embed CUAHSI JupyterHub directly into their class assignment page. (The potential for collaboration with HydroLearn is discussed more in [Section 6.3](#).)

This realization, combined with those discussed in [Section 6.2](#), led us to suspect that educators could be a key target user group for CUAHSI JupyterHub. In addition, we hypothesize that hydrologists still in school would be among those most open to adopting new tools. If young hydrologists were to adopt CUAHSI JupyterHub early on in their careers, then it might become an integral part of their toolkit, and they might continue using it beyond class.

Designing for students and educators would also help make the interface for CUAHSI JupyterHub as simple and easy-to-use as possible. Because such a design would assume no prior knowledge or skill set, CUAHSI JupyterHub would be accessible to hydrologists from all backgrounds and with all levels of experience.

# 5. Building a More Integrated Experience

In order to integrate the CUAHSI JupyterHub platform and the HydroShare data repository in a seamless and accessible manner, we decided to create a simple web app. It could easily be integrated with the existing CUAHSI JupyterHub system, and it would provide the desired functionality to the widest range of users with the least amount of development effort.

We determined that key functionality would include the ability to see all of a user's resources, to make opening a Jupyter notebook in a resource as easy as clicking a button, and to make syncing a resource's files between CUAHSI JupyterHub and HydroShare as simple as drag and drop. We intended for our app to make a clear distinction between what files are in CUAHSI JupyterHub and what files are in HydroShare.

## 5.1 User Interface and Experience Design Elements

Simplicity is paramount if CUAHSI JupyterHub is intended to be accessible to the widest audience possible. All it should take to get started using CUAHSI JupyterHub is basic computer literacy skills. This means that more technical interfaces such as the HydroShare API and the `hs_restclient` Python

module should be shielded from the user entirely, and the integration of the HydroShare data repository with CUAHSI JupyterHub should be as seamless as possible.

In order to minimize the mental burden of using the CUAHSI JupyterHub platform, we wanted to mimic existing prominent file management paradigms and models. We considered two general paradigms when designing our user experience. The first was one in which there was effectively one copy of the data in HydroShare. This would be similar to how Google Docs, Dropbox, or the like operate. Any action performed by the user, such as saving a Jupyter notebook, would be immediately reflected in HydroShare. Syncing would be automatic, and there would only ever appear to be one copy of the data. The other option we considered was one in which syncing had to be manually initiated by the user. Notebooks, data files, and the like could be edited in CUAHSI JupyterHub, and those changes would only be synced to HydroShare if the user performed an explicit action to transfer the data. We decided this latter option, in which syncing must be initiated manually, was the better choice. While it would require an extra step, it would reduce the potential for conflicts to arise if two people were working on the same resource at the same time.

Given the decision to require manual syncing of files between the CUAHSI JupyterHub platform and the HydroShare data repository, we needed to clearly communicate to the user the difference between the two locations. We settled on labeling the CUAHSI JupyterHub copy of the data “workspace files” in the app. This was meant to convey to the users that the files were able to be worked on, but that CUAHSI JupyterHub would be acting as a temporary location. (Unfortunately we came up with this term toward the end of the semester, so we were unable to get feedback on it from potential users.)

We were able to ask AGU Fall Meeting attendees what the most intuitive means of transferring files back and forth was, and the answer was clear: dragging and dropping. Nearly every person who was shown our mockups and asked how they would attempt to move files back and forth answered “dragging and dropping.” Therefore, we settled on building a user interface in which the user could drag and drop files between their workspace in the CUAHSI JupyterHub platform and the HydroShare data repository.

In summary, we decided our app would allow the user to manually transfer files between the JupyterHub platform and the HydroShare data repository in as simple a fashion as possible. It would leverage the familiar drag-and-drop file management paradigm, and it would clearly convey that JupyterHub and HydroShare were two different places in which files were stored. With this decided, the next step was to determine how we were going to build our app.

## 5.2 Web Technology Stack

Once we decided to make a web app to more seamlessly integrate CUAHSI JupyterHub and HydroShare, we considered several potential ways of doing this. The main criteria we considered when exploring options were the prominence of the technologies, how well supported they currently were, how well supported they likely would be in the future, and our own familiarity with the technologies involved.

First, we considered using the Jupyter notebook AppMode extension [4]. This would allow us to create a graphical web app of sorts using Python and Jupyter, which was attractive for several reasons. This would have allowed us to build off work a previous CUAHSI developer had done [5], and it would be a solution that CUAHSI developers, who already know Python, would easily be able to maintain. However, it would also have limited us quite significantly in what we could create. The building blocks for the interface appeared to be only basic buttons, text boxes, tables, and the like. Drag-and-drop was almost certainly unachievable. While the Jupyter notebook AppMode extension was simple, it was also very restrictive. We decided against using this technology..

With Jupyter AppMode notebooks off the table, we began to consider more common web development technologies. There were a multitude of options to consider, but we confined our search to those with which we and CUAHSI were already familiar.

We briefly considered Django [6] (which is what HydroShare uses). We decided against it because we did not feel it was a good fit for the kind of web app we were after. Django is a server-side web framework, meaning updating the web page (without using a separate JavaScript framework) would require fully refreshing the page. We felt this would create a slow and somewhat jarring user experience, if the page were constantly refreshing. Thus, we decided to steer away from Django in favor of a more JavaScript-driven web app.

One JavaScript framework we considered using was jQuery [7]. We concluded that jQuery would not be not the best choice for a modern web app. jQuery is one of the oldest JavaScript frameworks still in use today, and its original goal was to standardize web APIs across inconsistent browsers. It can be used to make a web page an interactive app, but building complex web applications was not its intended purpose. Many newer JavaScript frameworks have since been developed for the explicit purpose of creating web applications and offer a much better development experience. As far as we know, no major tech companies use jQuery to power a significant portion of their web stack today. We did recognize that CUAHSI developers have some familiarity with jQuery, but we felt that, if we were building something from scratch, using a modern JavaScript framework that was intended for app development was a better choice.

The final option we considered was to use the extremely popular React [8] JavaScript framework developed by Facebook. It powers much of Facebook's own technology stack and is also used by a multitude of other high-profile tech companies including Airbnb, Netflix, Slack, and Twitter. It is used to create single-page web applications with all page updates done client-side, thus resulting in a smooth and fast user experience. Two of our team members also had significant prior experience using React, meaning the time spent learning a new technology would be minimal. We recognized that CUAHSI developers did not necessarily have experience using React, and we brought up this concern with our liaisons. They told us using React would not be an issue, since they cared most about streamlining our development process. Thus, we decided to build our front-end using React.

In order to support a React front-end, we also needed to create a server-side component to interact with the files on CUAHSI JupyterHub. We decided to do this in Python and to utilize the `hs_restclient` API for all communication with HydroShare. We also realized that, on top of this, the backend could be written as a Jupyter server extension [9], and the entire app could be bundled as a Python package

and published to PyPi [10]. This would make installing our app as easy as running “pip install hydroshare\_jupyter\_sync.”

With a design outline in hand and a technology stack selected, we commenced work on the development of our app.

### 5.3 HydroShare Jupyter Sync Web App

Our web app consists of two pages: the first displays the user’s resources (Figure 4), and the second displays the contents of a particular resource (Figure 5). The general design of the pages was created to be relatively consistent with the appearance of JupyterHub.

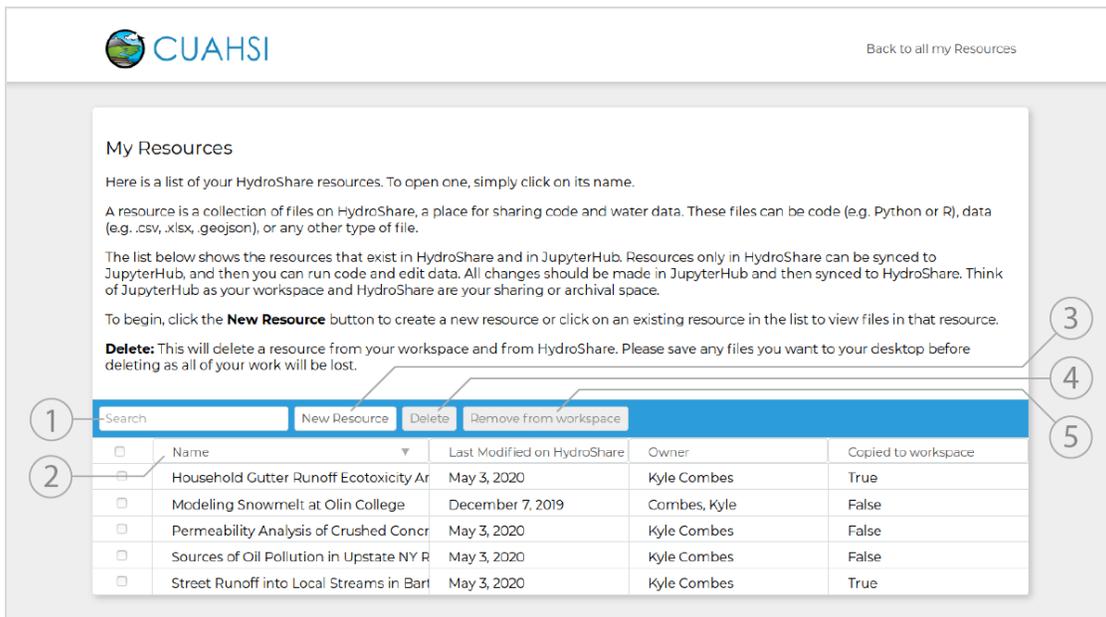


Figure 4: This page shows the user a list of their resources. This includes resources they have created or are collaborators on in HydroShare as well as any other resources the user has opened in CUAHSI JupyterHub in the past. The list of resources can be filtered by name (1) and sorted by column (2). The user can also create a new resource (3), delete a resource from HydroShare and CUAHSI JupyterHub (4), or just remove the copy of the resource in their CUAHSI JupyterHub workspace (5).

The resource list page (Figure 4) displays all of a user’s resources. This includes all of the resources the user has created (viewable under “My Resources” in HydroShare) as well as those created by others that the user has opened in CUAHSI JupyterHub in the past (via the “Open With” button on the resource page in HydroShare). This presents one unified view of all of the users’ resources and allows them to return to their work without the risk of accidental data loss (discussed in Section 3). The user can filter the resources by name (1) and sort by column (2). They can also create a new resource (3), which prompts them for a title and an abstract. Clicking the “Delete” button (4) will purge all resource files from the workspace and delete the resource from HydroShare. If the user just wants to remove the copy of a resource in their workspace, then they can select the resource and click on the “Remove from workspace” button (5). Clicking either of these buttons will display a confirmation dialog (explaining the consequences of the action) with a list of the selected resources before any action is taken. Clicking on the name of a resource opens the page shown in Figure 5.

CUAHSI Back to all my Resources

## Household Gutter Runoff Ecotoxicity Analysis

1 Creator: Kyle Combes Authors: Kyle Combes Created: May 3, 2020 Last Modified: May 3, 2020 Sharing Status: Private edit 2

Abstract  
An investigation into the effects of runoff from roof tiles into nearby ponds.

3 For help getting started with using Jupyter notebooks or using HydroShare data, please consult this [Helper Notebook](#).

4 Please note that data stored in your workspace is temporary and will be automatically deleted after 30 days of inactivity. When you are done working, please save your work in HydroShare by dragging and dropping your files and folders over. Visit [My Resources](#) to delete unused workspaces.

5 Workspace Files Jupyterhub

Filter	New	Upload	Delete
Name	Type	Size	
<input type="checkbox"/> Analysis	ipynb	74KB	
<input type="checkbox"/> Data	folder	48KB	
<input type="checkbox"/> Fish-lifespans	csv	11KB	
<input type="checkbox"/> Measurements	csv	36KB	
<input type="checkbox"/> README	md	267B	

6 HydroShare Files HYDROSHARE

Filter	Delete	Open in HydroShare
Name	Type	Size
<input type="checkbox"/> Analysis	ipynb	47KB
<input type="checkbox"/> Data	folder	48KB
<input type="checkbox"/> Fish-lifespans	csv	11KB
<input type="checkbox"/> Measurements	csv	36KB

9 README Display from: Workspace HydroShare

### Acid rain, roof tiles, and aquatic life

This (faux) analysis quantifies the impact of acid rain and roof tiles on nearby aquatic life.

#### Running the model

1. Open Analysis.ipynb
2. Run all of the cells

#### Contributors

- Kyle Combes ([website](#))

Figure 5: The resource details page shows a user resource metadata (1) with a link to edit the resource's sharing status in HydroShare (2). It also displays a welcome message with a link to a "getting started" notebook (3) and a message about the data retention practices of CUAHSI Compute (4). The user can also manage files within their workspace (5) and in HydroShare (6) and transfer files between the two by dragging and dropping. The "Open with HydroShare" button (5) will open the resource in HydroShare. If the resource contains a readme, that is displayed at the bottom of the page (9).

At the top of the page shown in Figure 5 is a list of resource metadata (1). Among the metadata displayed is the sharing status on HydroShare as well as a button to edit the sharing settings (2). We did not want to recreate the privacy and collaborators editor in HydroShare, so clicking on this button simply redirects the user to the resource page on HydroShare.

Below the metadata is an optional welcome message directing the user to a "getting started" notebook (3). The path to this notebook is specified in the config file, and this message will only be

displayed if the config file specifies the path to this notebook. Clicking the “Helper Notebook” link will open the notebook.

Below the welcome message is an optional message informing the user of the data retention practices for the platform they are using (4). If the user is using CUAHSI JupyterHub on Binder, which is a service that temporarily hosts CUAHSI JupyterHub just for them, then this message should notify them that all workspace files will be lost after the user leaves. If the user is using CUAHSI JupyterHub running on the RENCI computing cluster, where information persists but storage is limited, it could ask them to remove the workspace copy of any resources they are no longer working on. The exact message can be specified in the config file. If no message is specified, then this notice will not appear.

The two cards displayed side-by-side (5 and 6) allow the user to sync files between CUAHSI JupyterHub (5) and HydroShare (6) by dragging and dropping files between the two. They can also reorganize files within CUAHSI JupyterHub and HydroShare by dragging and dropping within their respective pane.

To open a file in Jupyter, the user can simply click on a file in the Workspace Files pane (5). New files and folders can also be created in the Workspace Files pane, and the user can also upload a file from their local computer to their Workspace Files. Once a file or folder has been created or uploaded, the user can copy them to HydroShare via dragging and dropping. Bulk deletion of files and folders is also available for both locations using the checkboxes and Delete buttons. Clicking the “Open in HydroShare” button (8) will open the resource page in HydroShare.

While the functionality for collapsing folders does exist, it is currently disabled. Enabling the collapsing of folders results in the breaking of the drag and drop functionality, for a reason that is not currently known. More information on this issue can be found in Issue #27 on our GitHub repository [\[13\]](#).

Below the file manager is a rendering of the resource’s README.md (9). If such a file exists, the Markdown is rendered and displayed in its entirety. This makes it easy for resource authors to include run instructions and have them automatically displayed to anyone who opens their resource.

Since the page shown in Figure 5 gives users an orientation to CUAHSI JupyterHub and could display a readme containing instructions to run models in a resource, we recommend that users be redirected to this page after clicking “Open With” in HydroShare.

The interface we have created lays the foundation for a streamlined and easier-to-use integration of the JupyterHub platform and the HydroShare data repository. Our liaisons have expressed nothing but excitement toward our app, and it is now in being deployed for preliminary testing.

## 6. Next Steps

Our resource file sync web app lays a solid foundation for improving the integration of the CUAHSI JupyterHub platform with the HydroShare data repository. Users can now manage their files in a graphical manner, without needing to use the `hs_restclient` Python module if they so choose. Even

with the progress we have made, there is much that can still be done to improve the user experience further. We have captured below some of our ideas we were unable to implement. They are listed in the order in which we perceive them to be most important or impactful.

## 6.1 Remaining Issues

We have documented a number of remaining bugs and minor recommendations in the Issues for our GitHub repository [\[12\]](#). Developers should refer to those for the most up-to-date information on bugs and feature recommendations.

## 6.2 Educational Resources

The most salient insight we gained when talking to hydrologists revolved around their experiences learning to write code. When asked, all of the hydrologists we talked to lamented their experiences learning to code. They felt existing resources missed the mark in terms of providing an attractive introduction to coding for hydrologists.

Googling and learning from peers were the most common means of learning to code, and the languages and libraries used were largely dictated by what new hydrologists' research groups were already using. Extensive Googling in order to learn fundamental coding concepts often resulted in significant amounts of hair-pulling and frustration. A substantial portion of the time was being dedicated to finding the right resource, not to actually learning the material.

What introductory coding resources do exist are often minimally relevant to hydrologists. Computer Science classes covering low-level topics like memory management and sorting algorithms provide limited value given the time and energy commitments needed for the classes. The only introductory coding resource for scientists specifically that we came across during our exploration was Software Carpentry [\[13\]](#), which nobody we interviewed had ever used. The materials provided on the website were also extremely limited and intended to be paired with an in-person guided workshop.

Given the identified lack of introductory coding resources relevant to hydrologists, a consolidated set of resources introducing hydrologists to the fundamentals of computing would be immensely helpful. This could include how to get started with Python, how to work with data programmatically, and what relevant software libraries exist. These would ideally be organized into a logical progression of coding skills development, in a way that a "Computing for Hydrologists 101" class might structure topics. Even short of developing an entire curriculum, we believe that a curated set of existing resources in a format accessible to new hydrologists would create substantial value.

## 6.3 Collaborating with HydroLearn

In our user interviews (documented above in Section 4.2.2), we learned that several of the people behind the HydroLearn platform are already using the CUAHSI JupyterHub platform in their HydroLearn courses. The web app we have created will hopefully make it easier for them and their students to edit code in Jupyter notebooks. That said, there is still much work in this space that can be

done to improve the workflow for students and educators and to make them continue using the HydroShare data repository and the CUAHSI JupyterHub platform. For example, some educators expressed the desire to embed the Jupyter notebooks directly into their website. Additionally, they would greatly appreciate the functionality described below in [Section 6.4](#) to allow for easy transfer of data between students' and teachers' resources.

## 6.4 Sync Between Resources

One use case for our HydroShare Jupyter Sync app could be to transfer files between HydroShare resources. This would allow users to copy files both between their own resources and into their own resources from others' resources. The latter functionality could prove particularly useful to educators, who often use one resource to distribute homework sets and then ask students to submit their work in their own resource. Enabling students to copy files from their instructor's resource into their own personal resource could be very helpful.

Enabling such support in the web app would not be particularly difficult. We have documented our thoughts on how this might be accomplished in Issue #28 on our GitHub repository [[14](#)].

## 6.5 Collaboration Tools & Version Control

When talking to hydrologists, we learned that some of those who use computing also use a version control system such as Git. We do not have a good sense of what fraction of hydrologists use versioning or what the demand for version control in CUAHSI JupyterHub might be, since it was not a question we regularly asked during our interviews. However, we do believe that code versioning is an important component of good software development practices. Thus, incorporating a form of version control into CUAHSI JupyterHub and encouraging its use may present significant benefit to hydrologists.

## 6.6 Access CUAHSI JupyterHub Without Using HydroShare

While the CUAHSI JupyterHub platform is currently only intended to be accessed via the HydroShare data repository website, we thought it could be valuable to make another entry point for those who aren't familiar with HydroShare. This could help tailor to students and other first-time users who may just be looking for a place to run Python or R code. Instead of requiring users to navigate HydroShare, a new user could just visit [jupyter.cuahsi.org](http://jupyter.cuahsi.org), click a "sign up" button, be redirected to the sign-up page on [hydroshare.org](http://hydroshare.org) (since they need a HydroShare account to use CUAHSI JupyterHub), and then be redirected directly to the page displayed in Figure 4. If such a second entry point were created, visitors should be greeted with a landing page. Figure 6 is a draft of what we thought such a landing page might look like.


Login



## Welcome to CUAHSI JupyterHub.

Analyze data, build models & more using Python or R right in your browser. For free.



**Click & Go**

No more setup struggles. Everything you need is ready to go. Just type your code and click Run.



**Fast**

Your code runs in the cloud, so it doesn't matter if your computer is slow.



**Connected**

Easily download data and models from HydroShare. Share your code when you're ready.

---

### Getting Started

As part of the HydroShare ecosystem, CUAHSI JupyterHub uses the same login system as HydroShare. Simply login with your existing HydroShare account or create a new account for free if you don't have one.

Login with HydroShare
Create a HydroShare account

---

### Empowering You to Do More



Designed for hydrologists, our online programming environment is built off the popular JupyterHub platform. It's the easiest way to start coding.



Our Jupyter setup integrates seamlessly with HydroShare, allowing you to easily load data and models others have published.

Figure 6: It might be valuable to design a second entry point into CUAHSI JupyterHub (other than via HydroShare) so that new users need only familiarize themselves with one system. This might also better tailor to users who are merely looking to run Python or R code easily. Instead of starting in HydroShare, users could simply go to [jupyter.cuahsi.org](http://jupyter.cuahsi.org) and be presented with a landing page such as this to welcome them and introduce the platform.

## 7. Conclusion

CUAHSI JupyterHub currently provides hydrologists with a free JupyterHub-based computing environment, but its user interface and user experience have yet to be fully refined. Through our user research, we determined that experienced hydrologists are predominantly content with their current compute solutions, whereas less experienced hydrologists are less satisfied with the current offerings. Some educators are also using Jupyter in their classes, though not all are able to do so easily. Given these insights, we recommend educators and their students be considered the primary users when refining the user experience of CUAHSI JupyterHub.

In order to streamline the integration of CUAHSI JupyterHub with HydroShare, we built a web app using React and Python. This app, called CUAHSI JupyterHub Sync, allows users to easily manipulate files in CUAHSI JupyterHub and synchronize them with HydroShare. It reduces the number of clicks required to synchronize data from at least six performed across three pages down to a single click. This significantly reduces the time and effort needed to synchronize data between CUAHSI JupyterHub and HydroShare.

We also documented the improvements we thought of but did not have time to make in this report and on GitHub. The most impactful would be adding the ability to sync between resources and collaborating with HydroLearn.

Additionally, creating a curated list of existing resources for learning Python would also address the major gap we identified in introductory educational resources. The most salient insight we gleaned from our research was a perceived lack of educational resources for hydrologists to learn relevant coding concepts in the context of hydrology. Filling this hole by pointing self-learners in the right direction would bring immense value to the hydrology community.

Our user research and software development work are a substantial step forward for CUAHSI JupyterHub. With a little more polish, CUAHSI JupyterHub will reach its full potential to empower the hydrology community by making powerful, accessible computing freely available to all.

# References

1. Consortium of Universities for the Advancement of Hydrologic Science, Inc. "Python client for the <https://www.hydroshare.org> REST API." GitHub. [https://github.com/hydroshare/hs\\_restclient](https://github.com/hydroshare/hs_restclient).
2. Consortium of Universities for the Advancement of Hydrologic Science, Inc. "Using the API." HydroShare. <https://www.hydroshare.org/hsapi/>. Accessed 6 Oct. 2019.
3. "Step 1. Is Your Project Considered Research?" University of Washington. <https://www.washington.edu/research/hsd/do-i-need-irb-review/is-your-project-considered-research/>. Accessed 2 May 2020.
4. Schütt, Ole. "Appmode." GitHub, <https://github.com/oschuett/appmode>. Access 2 May 2020.
5. Hunt, Martin. (2019). "Resource Index." HydroShare, <http://www.hydroshare.org/resource/92ade040a75647ed8c9be452a86d90f5>.
6. Django Software Foundation. "Django: The Web Framework for Perfectionists with Deadlines." <https://www.djangoproject.com/>. Accessed 2 May 2020.
7. "jQuery: Write Less, Do More." <https://jquery.com/>. Accessed 2 May 2020.
8. "React: A JavaScript Library for Building User Interfaces." <https://reactjs.org/>. Accessed 2 May 2020.
9. Project Jupyter. "Distributing Jupyter Extensions as Python Packages - Jupyter Notebook 6.0.2 documentation." <https://jupyter-notebook.readthedocs.io/en/stable/examples/Notebook>. Accessed 16 Oct. 2019.
10. Python Software Foundation. "Find, Install and Publish Python Packages with the Python Package Index." PyPI, <https://pypi.org/>. Accessed 2 May 2020.
11. GitHub HydroShare JupyterHub Sync Issue #27: Collapsible Folder and Drag and Drop. [https://github.com/hydroshare/hydroshare\\_jupyter\\_sync/issues/27](https://github.com/hydroshare/hydroshare_jupyter_sync/issues/27). Accessed 3 May 2020.
12. GitHub HydroShare Jupyter Sync Issues. [https://github.com/hydroshare/hydroshare\\_jupyter\\_sync/issues/](https://github.com/hydroshare/hydroshare_jupyter_sync/issues/). Accessed 3 May 2020.
13. Software Carpentry. <https://software-carpentry.org/>. Accessed 13 Nov. 2019.
14. GitHub HydroShare JupyterHub Sync Issue #28: Sync between resources. [https://github.com/hydroshare/hydroshare\\_jupyter\\_sync/issues/28](https://github.com/hydroshare/hydroshare_jupyter_sync/issues/28). Accessed 3 May 2020.

# Appendices

## A. Initial User Research

We initially reached out to a wide variety of hydrologists and interviewed them over the phone prior to conversations we had at the AGU Fall Meeting. The interview protocol we used is included below.

### INTERVIEW GOALS

- What are their major needs?
- Gain an understanding of:
  - What frustrations they face
  - What tools they might be lacking
- Might they be interested in follow-up interviews?

### QUESTIONS

#### Background & Values (5 min)

- How do you prefer to be addressed?
- Quick background (4 min)
  - What is your background?
  - What do you do now?
  - What data are you interested in?
- Values
  - What is your overarching goal with your research?
    - Ex: “I want to get published”, “I want to learn”, “I want to make people safe”
  - What interests you about hydrology?
  - What defines success for you?

#### Existing resources

Have you used any of the following introductory coding resources? If not, what do you think when you look at these pages? Are you inclined to try any of them? Why or why not?

- **Software Carpentry**
  - <http://swcarpentry.github.io/python-novice-inflammation/>
  - <http://swcarpentry.github.io/python-novice-inflammation/02-loop/index.html>
  - <http://swcarpentry.github.io/python-novice-inflammation/06-func/index.html>
- **Python for Data Science** <https://www.edx.org/course/python-for-data-science-3>
- **Using Python for Research** <https://www.edx.org/course/using-python-for-research-2>

#### Programming tools

- What environment do you use? (Jupyter on your computer, Jupyter hosted by someone else, text editor, IDE)
- What libraries do you use a lot?
- What APIs or data sources do you use?
- How are you learning these different tools?

- What is your preferred learning style? (Lectures or figure it out yourself)

### Focusing on our space

- Use & Manipulate section:
  - Is there a common question you seek to answer when working with this type of data?

## B. User Research at the 2019 AGU Fall Meeting

We attended the American Geophysical Union (AGU) 2019 Fall Meeting in San Francisco during the last week of the semester. We did so in order to confirm our research from earlier in the semester, get feedback on the web app we are designing, and learn more about users' workflows and how JupyterHub could fit into them. In particular, we aimed to validate our terminology choices in our web app to ensure they are intuitive for hydrologists, and to make sure we are clearly communicating the functionalities and benefits of CUAHSI JupyterHub. We discuss the activities we conducted and the preliminary findings from those activities in this section.

### B.1 Moving a Resource Activity

The goal of this activity was to confirm the language we used to synchronize files between JupyterHub and HydroShare. We created a piece of paper (Figure 7) to represent a space for files on JupyterHub and another to represent files on HydroShare (these are pictured below). We then asked the participant to pretend a sticky note was a file, and move that sticky note from the JupyterHub paper to the HydroShare paper. As they did so, we asked them to give us a word for the action they were performing. We did this to determine what terminology our app should use.

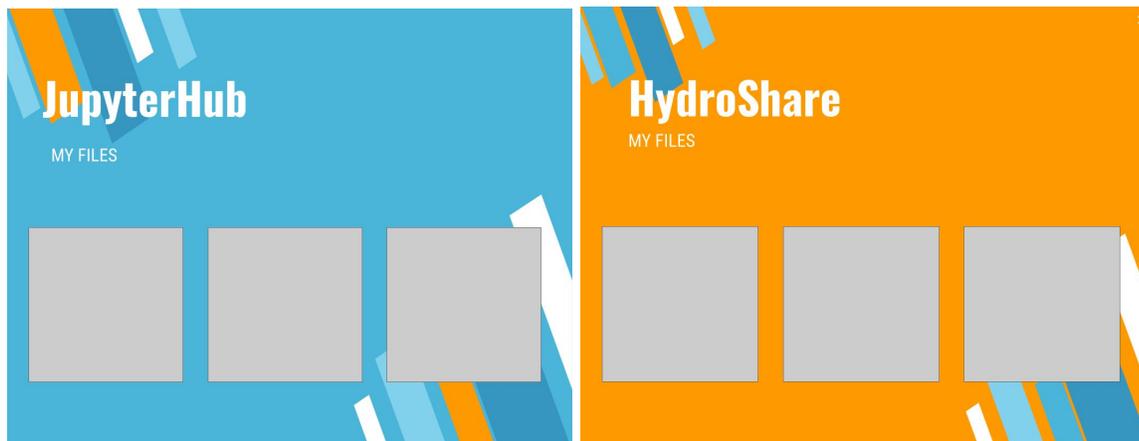


Figure 7: We used these two mats when asking users to move files back and forth while sharing the terminology they would use to describe their actions.

### B.2 Workflow and Mood Card Activity

We created a deck of cards with words that participants could use to create their workflow from the beginning of a project to the end, including emotions they could assign to each part of the workflow. The goal of this activity was to learn about typical workflows and pain points within those workflows.



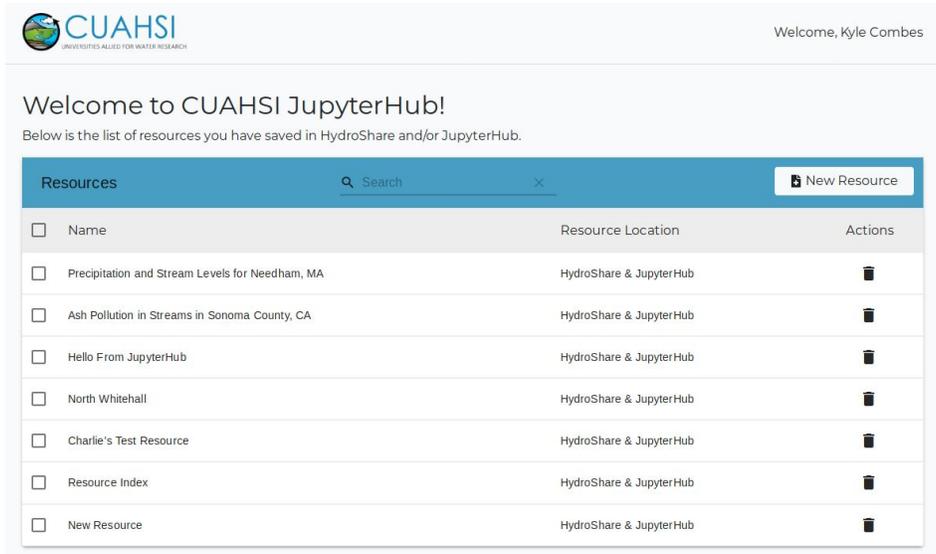


Figure 9: A screenshot of a partly functional but rough prototype of our app. We asked AGU Fall Meeting attendees for their feedback on this page.

## Resource Page

We asked attendees the following questions about Figure 10:

- What information is this page providing you?
- What actions can you take on this page?
- What action would you take to sync your Jupyter files with HydroShare?
- What are the two views (JupyterHub and HydroShare columns)?
- Any other thoughts?

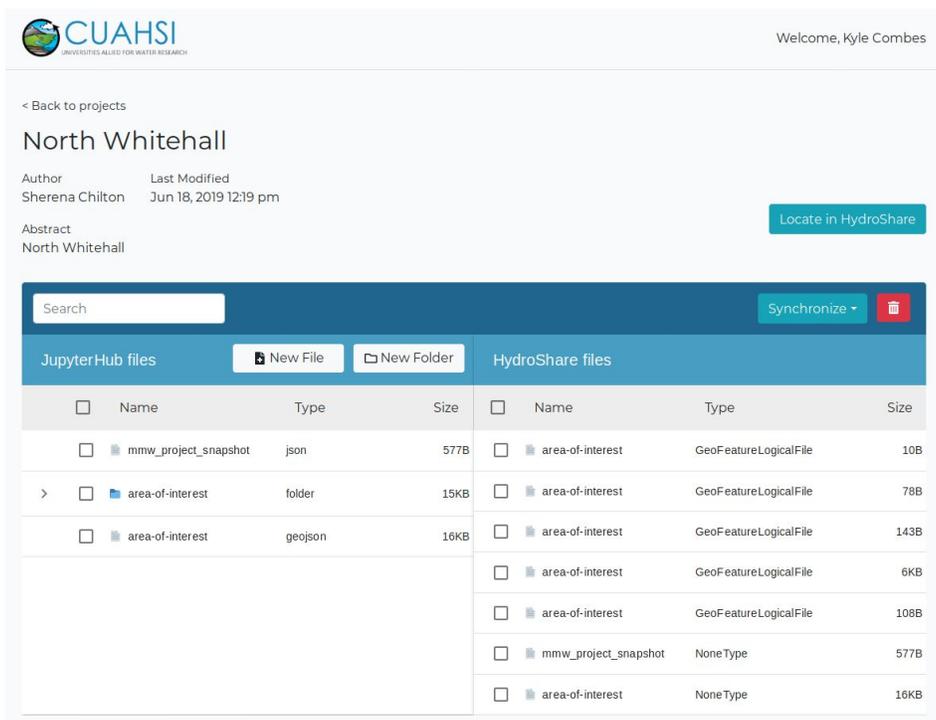


Figure 10: A screenshot of our partly functional app prototype. We asked AGU Fall Meeting attendees for feedback on this page.

## B.4 Survey

We created a Google survey to get quick information from participants. It asked questions about the participant's background in hydrology and relationship to computing tools or code sharing tools, and contact information.

## B.5 General Interviews and Discussions

Given that we were working in a public space at a conference, we had many discussions with people that weren't guided by any activity.

## C. Directory Exists Prompt

Users currently see the following prompt when they click the "Open with CUAHSI JupyterHub" button for a resource that they have opened with CUAHSI JupyterHub in the past.

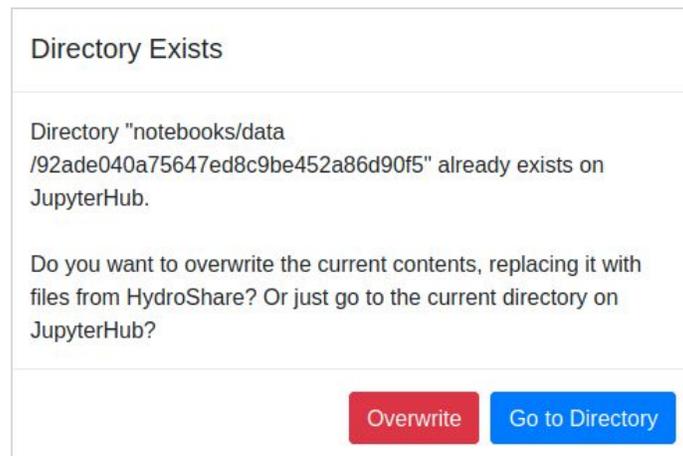


Figure 11: The message that CUAHSI JupyterHub currently displays users when they try to open a resource from HydroShare after having already opened it once before.