

Scientific Data Annotation and Dissemination: Using the ‘Ike Wai Gateway to Manage Research Data

SEAN B. CLEVELAND, University of Hawai‘i

JENNIFER GEIS, University of Hawai‘i

GWEN A. JACOBS CLEVELAND, University of Hawai‘i

Granting agencies invest millions of dollars on the generation and analysis of data, making these products extremely valuable. However, without sufficient annotation of the methods used to collect and analyze the data, the ability to reproduce and reuse those products suffers. This lack of assurance of the quality and credibility of the data at the different stages in the research process essentially wastes much of the investment of time and funding and fails to drive research forward to the level of potential possible if everything was effectively annotated and disseminated to the wider research community. In order to address this issue for the Hawai‘i Established Program to Stimulate Competitive Research (EPSCoR) project, a water science gateway was developed at the University of Hawai‘i (UH), called the ‘Ike Wai Gateway. In Hawaiian, ‘Ike means knowledge and Wai means water. The gateway supports research in hydrology and water management by providing tools to address questions of water sustainability in Hawai‘i. The gateway provides a framework for data acquisition, analysis, model integration, and display of data products. The gateway is intended to complement and integrate with the capabilities of the Consortium of Universities for the Advancement of Hydrologic Science’s (CUAHSI) Hydroshare by providing sound data and metadata management capabilities for multi-domain field observations, analytical lab actions, and modeling outputs. Functionality provided by the gateway is supported by a subset of the CUAHSI’s Observations Data Model (ODM) delivered as centralized web based user interfaces and APIs supporting multi-domain data management, computation, analysis, and visualization tools to support reproducible science, modeling, data discovery, and decision support for the Hawai‘i EPSCoR ‘Ike Wai research team and wider Hawai‘i hydrology community. By leveraging the Tapis platform, UH has constructed a gateway that ties data and advanced computing resources together to support diverse research domains including microbiology, geochemistry, geophysics, economics, and humanities, coupled with computational and modeling workflows delivered in a user friendly web interface with workflows for effectively annotating the project data and products. Disseminating results for the ‘Ike Wai project through the ‘Ike Wai data gateway and Hydroshare makes the research products accessible and reusable.

CCS Concepts: • **Information systems** → **Data management systems**.

Additional Key Words and Phrases: datasets, annotation, data, dissemination, discovery, FAIR

ACM Reference Format:

Sean B. Cleveland, Jennifer Geis, and Gwen A. Jacobs Cleveland. 2020. Scientific Data Annotation and Dissemination: Using the ‘Ike Wai Gateway to Manage Research Data. In *Practice and Experience in Advanced Research Computing (PEARC ’20)*, July 26–30, 2020, Portland, OR, USA. ACM, New York, NY, USA, 16 pages. <https://doi.org/10.1145/3311790.3396661>

1 INTRODUCTION

Granting agencies invest millions of dollars on the generation and analysis of data, making these products extremely valuable. However, without sufficient annotation of the methods used to collect and analyze the data, the ability to

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

© 2020 Copyright held by the owner/author(s).

Manuscript submitted to ACM

reproduce and reuse those products suffers. This lack of assurance of quality and credibility of the data at the different stages in the research process essentially wastes much of the investment of time and funding and fails to drive research forward to the level of greatest potential if everything was effectively annotated and disseminated to the wider research community.

The challenges of managing scientific data are significant, and over the years they have typically fallen in the hands of investigators. There exist significant obstacles in workflows supported by cyberinfrastructures; from operation and field deployment of sensors to data streams – to data management – to data analysis – to the use of integration tools. These multifaceted obstacles involve hardware, middleware, and software. However, significant work and progress has been made to tackle the challenges of managing these workflows, discovering data, storing data, and publishing scientific data in architectures that are conducive to ease-of-use, dissemination, documentation, and research for scientists. Principal investigators, researchers, managers, and scientists alike need the ability to easily access (and possibly integrate) information that is housed in distinct geographical and distributed sites. Additionally, such information is very likely to be stored in different formats and disseminated using a diverse range of communication protocols.

Data should be managed to:

- Maximize the effective use and value of data and information assets.
- Continually improve the quality including data accuracy, integrity, integration, timeliness of data capture and presentation, relevance, and usefulness.
- Ensure appropriate use of data and information to facilitate data sharing.
- Ensure sustainability and accessibility in the long term for re-use in science.

If data is well-organized, documented, preserved, accessible, and verified as to accuracy and validity, then the result is high quality data that is easy to share and re-use in science and yields a return on the investment of grant dollars.

In order to address the array of needs of statewide community stakeholders and the research aims of the Hawai'i EPSCoR project, a new open source water science gateway was developed at the University of Hawai'i (UH), called the 'Ike Wai gateway [1]. In Hawaiian, 'Ike means knowledge and Wai means water. The 'Ike Wai gateway aims to provide centralized web based user interfaces and application programming interfaces (APIs) supporting multi-domain data management, computation, analysis, and visualization tools to support reproducible science, modeling, data discovery, and decision support for the Hawai'i EPSCoR 'Ike Wai research team, water managers, decision makers, and the wider Hawai'i water community in order to answer questions related to water sustainability.

While the gateway is the primary tool for 'Ike Wai researchers to store, annotate, process, and analyze their data, the general public can't access the Gateway, so the data is out of their reach. To remediate this, we've designed two processes through which 'Ike Wai researchers can share their data with local stakeholders and the broader research community. One process "pushes" the metadata and links to the files to Hydroshare [16] and the other makes the annotations and data available in the 'Ike Wai data gateway, (<http://ikewai.org>). The 'Ike Wai data gateway is a web application in which all annotated data products that have been pushed to it are freely available to the public along with tools to help users find and understand the data.

This paper will describe the 'Ike Wai gateway's design and use of these data annotation, dissemination, and discovery features.

1.1 Background

1.1.1 'Ike Wai. The current Hawai'i EPSCoR Track 1 project is named 'Ike Wai, has a mission to ensure Hawaii's future water security through an integrated program of research, education, community engagement, and decision support. Its vision is to promote water resource management in Hawai'i that is sustainable, responsible, and data driven. Scientific, cultural, and social dimensions to the problem of water security are integrated in a transparent, stakeholder-driven, and rigorous water research enterprise. In order to support this mission a robust cyberinfrastructure solution such as a science gateway was required.

Science gateways and science gateway frameworks have become popular for providing research infrastructure and are being supported by organizations such as the Extreme Science and Engineering Discovery Environment (XSEDE) [3] and the Science Gateways Community Institute (SGCI) [17] [8] [9]. Several efforts to provide 'as a service' gateways are: WS-PGRADE/gUSE project [18], the HUBzero project [12], the Apache Airavata [4] [7], Galaxy [2] and Globus [5] has a cloud-style model for managing data transfers and sharing data. Existing hydrology gateways and tools such as the CUAHSI Hydrological Information System (HIS) [15], Hydroshare [16], and Virtual Observatory and Ecological Informatics System (VOEIS) [11] provide exceptional server side support, data entry, and quality control client tools. However, existing water data gateways are lacking tools for managing and validating multi-domain metadata, lack of integration for multiple data stores, as well as non-integration of compute resources, or flexibility in access and selection of those resources.

To address these challenges Tapis [14], built off Agave [6], was chosen as the framework to build the 'Ike Wai gateway due the wide range of supported functionality that enable supporting the above needs and UH's partnership with the Texas Advanced Compute Center (TACC) that support it. Tapis is an open source, scientific API middleware for hybrid cloud computing and data management that is powering a number of current community science gateways including CyVerse [13] [10], which provides life scientists with powerful computational infrastructure to handle huge datasets and complex analyses that enable data-driven discovery.

The 'Ike Wai gateway supports research in hydrology and water management and provides tools to address questions of water sustainability in Hawai'i. The gateway provides centralized web based user interfaces and APIs supporting multi-domain data management, computation, analysis, and visualization tools to support reproducible science, modeling, data discovery, and decision support for the 'Ike Wai research team and wider Hawai'i hydrology community. By leveraging the Tapis framework [14] UH has constructed a gateway that ties data management and advanced computing resources together to support diverse research domains including microbiology, geochemistry, geophysics, economics, and humanities.

1.2 FAIR data support

The 'Ike Wai gateway support FAIR data principles guideline aiming to enhance the reusability of data and metadata. The FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. The guidelines and support are outlined below.

To be Findable [19]:

- F1: (meta)data are assigned a globally unique and persistent identifier
- F2: data are described with rich metadata
- F3: metadata clearly and explicitly include the identifier of the data it describes
- F4: (meta)data are registered or indexed in a searchable resource

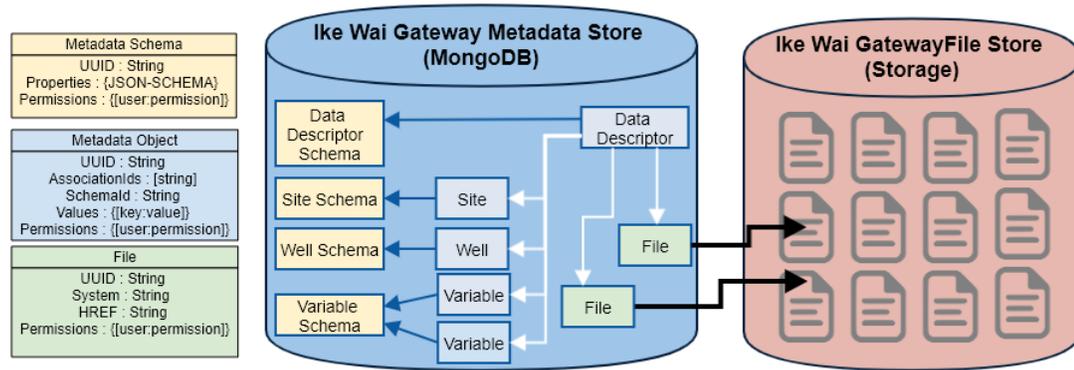


Fig. 1. Diagram of the 'Ike Wai gateway's Metadata and object design and file annotation.

The 'Ike Wai gateway supports Findable data by utilizing both unique identifiers in the metadata, URLs of data/file resources and DOIs through Hydroshare to address F1 (Figure 1,2). Further the use of controlled vocabularies and minimal metadata standards address F2 and F3. Metadata is stored and indexed for full-text and geospatial search to address F4.

To be Accessible:

- A1: (meta)data are retrievable by their identifier using a standardized communications protocol
- A1.1 : the protocol is open, free, and universally implementable
- A1.2: the protocol allows for an authentication and authorization procedure, where necessary
- A2: metadata are accessible, even when the data are no longer available

To address A1 the gateway requires all requests for metadata or file resources to be authenticated and authorized over HTTPS. A2 aims to be supported by the sustainability of the gateway as well as the metadata existing in Hydroshare.

To be Interoperable:

- I1: (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2: (meta)data use vocabularies that follow FAIR principles
- I3: (meta)data include qualified references to other (meta)data

I1,2 and 3 are support through the use of JSON metadata documents that have validated schema that leverage Dublin Core minimal metadata standards and CUAHSI controlled vocabularies.

To be Reusable:

- R1: meta(data) are richly described with a plurality of accurate and relevant attributes
- R1.1: (meta)data are released with a clear and accessible data usage license
- R1.2: (meta)data are associated with detailed provenance
- R1.3: (meta)data meet domain-relevant community standards

The gateway's metadata documents, such as the "Data Descriptor", described in more detail later, support reusability by requiring all relevant attributes, licensing and community standards. Further, the integration of a data manager in the loop ensures compliance.

2 IMPLEMENTATION

2.1 Data storage

The 'Ike Wai gateway allows researchers to manage data in a user space deemed "My Data". Within "My Data" the interface appears as a navigable directory structure allowing users to create, delete, and move folders for organization and move up and down the directory structure (Figure 3). Within the folders, researcher can upload, download, copy, move, rename, and preview file contents in supported formats such as images, txt, and pdf. Two additional storage areas are also available for the researchers. The "Team Data Preview" allows the researchers read only access to the data from the rest of the 'Ike Wai team for preliminary sharing and review. The "Annotated Team Data" is the area for files that have been annotated and quality controlled for usage by other team members and internal stakeholders.

2.2 Metadata

The gateway utilizes the Tapis Meta APIs which support storage and management of metadata as JSON documents that conform to defined schema for validation but also allows flexibility for expansion of schema definitions (Figure 1). The set of metadata objects or schema in the 'Ike Wai gateway consist of a Data Descriptor object, a Variable object, and Location objects such as Wells and Sites. The Data Descriptor object contains fields that contain the Dublin Core minimal representative fields (i.e., title, author, source, publisher, rights, description, etc.). Variable objects describe information about the data measured or computed such as its type, unit, medium (etc), and method description. Location objects are used to add spatial information and additional descriptions of where research activities occurred.

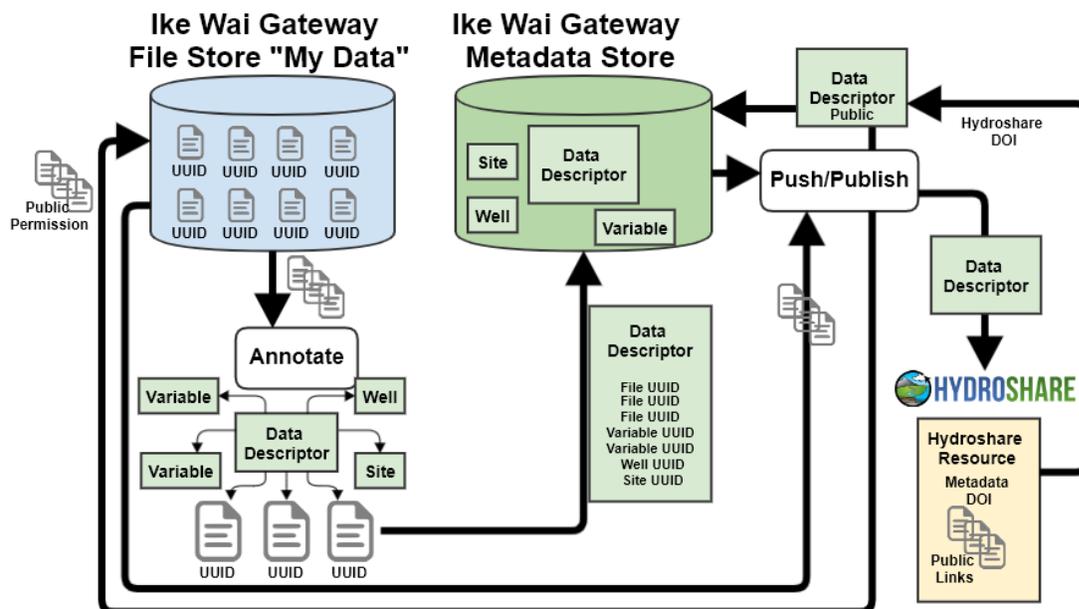


Fig. 2. Diagram of the 'Ike Wai gateway's annotation workflow and dissemination.

2.3 Data Annotation

The primary means through which data annotation is accomplished in the gateway is through the “Data Descriptor” (Figure 2,6). Within the ‘Ike Wai Gateway the Data Descriptor is presented as a web form containing fields that are deemed necessary to allow a user to document their data in such a way that it enables its usefulness to other users. The actions are organized around the context of data being in files within a user’s “My Data” space. To annotate a file, the user clicks on the “Annotate File” menu item in the “Actions” menu (Figure 3).

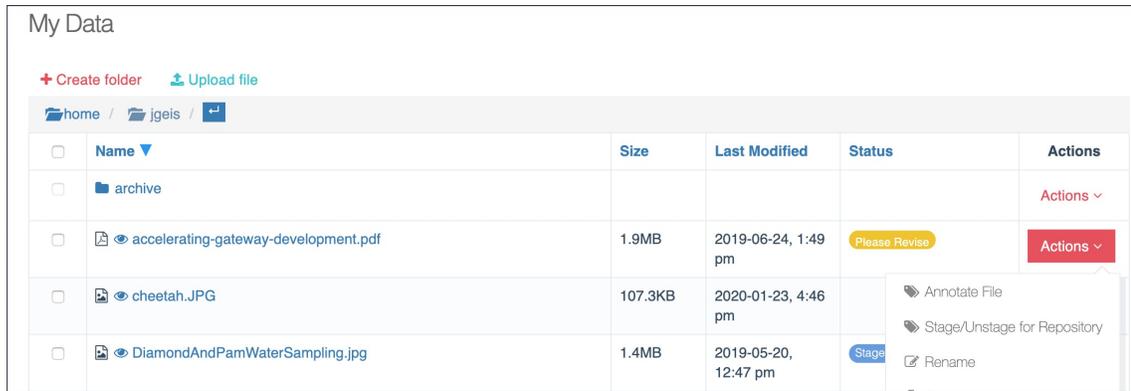


Fig. 3. Screenshot of ‘Ike Wai gateway’s “My Data” page.

This brings the user to the annotations page (Figure 4). Any existing Data Descriptors that have been associated with this file will be displayed at the top.

Viewing Annotations For:

 accelerating-gateway-development.pdf

Current Annotations (Data Descriptors associated with the above file)

Title	Summary	Actions
 Conference paper for Gateways 2019	Gateways 2019 conference paper submission discussing the 'Ike Wai Gateway.	Remove Association Clone + Edit

Fig. 4. Screenshot of existing associated Data Descriptors in the 'Ike Wai gateway.

The bottom includes an option to create new Data Descriptors, it also lists existing Data Descriptors that the user may choose to associate with their file or clone and edit.

Data Descriptors available for association:

[Q Show Search](#) [+ Create New Data Descriptor](#)

Total Objects: 20

Filter Current Local Search Results		
Diamond Tachera sampling rain water at UHH CC in August 2018	Diamond Tachera, Graduate Research Assistant, Dept. of Earth Sciences, SOEST, sampling rain water at the UH Hawai'i Community College Pālanui Campus in August 2018. Photo taken by Dr. Kiana Frank, Assistant Professor, PBRC	Associate Clone
Diamond and Sheree sampling rainwater on Mauna Kea	A photo of Diamond and Sheree sampling rainwater on Mauna Kea.	Associate Clone
Diamond and Pam collecting groundwater from a NELHA monitoring	Diamond Tachera, Graduate Research Assistant, Dept. of Earth Sciences, SOEST and Pamela Madden, Water Quality Specialist,	Associate Clone

Fig. 5. Screenshot of available Data Descriptors in the 'Ike Wai gateway.

The 'Ike Wai Data Descriptor allows the researcher to capture the following information: Title, Author, Data State, Contributing Organizations, Keywords, Start Date, End Date, Data Type, Format, Summary, Variables, and Location (Figure 6). These are the fields that apply to most of the data generated in the 'Ike Wai project and correspond to the minimal set of information required to annotate data. There are also a few special sections for domains that had different needs. For example, one domain in the project worked on translating Hawaiian Language newspaper articles into English. The gateway provides a special section in which researchers can indicate the newspaper, author and translator of each translated article. To better support hydrology, the gateway also includes a section with fields that make the Data Descriptor better align with Hydroshare's resource data model to support a researchers ability to generate data annotations in Hydroshare and obtain a Digital Object Identifier (DOI) with the click of a few buttons. These fields are Contributing Persons, Related Resources, and License Rights.

A single Data Descriptor can be associated with multiple files and a single file can be associated with multiple Data Descriptors. Note that when a researcher pushes data to Hydroshare or Ikewai.org, it is done via a Data Descriptor instead of through a file. This is because if the push process started upon selecting a file there could be multiple data descriptors available causing confusion and more complexity in the workflow.

2.4 Disseminating Data Products

We call the process of making the data available on Hydroshare or Ikewai.org "pushing." The user clicks a menu item to start the process which copies the metadata, and generates permanent publicly accessible links to the data files, over to the intended target.

Required

Title: ?

Author(s): ?

+Create New Author

Contributor - Agency: ?

+Add Contributing Organizations

Fields for Hydroshare

Fields for Hawaiian Language group only

Subject/Keywords/Search terms: ?

Data Collection Start Date: ?

📅

Data Collection End Date: ?

📅

Data State: ?

Raw Processed Preliminary product/result Final product/result Unknown

Can this data be made available to the general public?: ?

Public Private

Data Type: ?

Format: ?

Summary: ?

Fig. 6. Screenshot of a new Data Descriptor in the 'Ike Wai gateway.

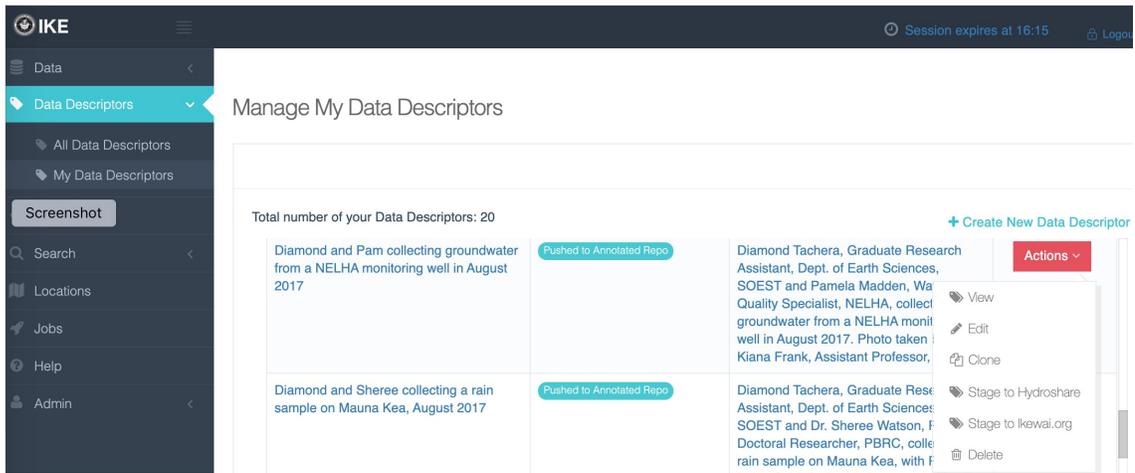


Fig. 7. Screenshot of the 'Ike Wai gateway Data Descriptor management page showing the menu for a selected data descriptor.

When the user clicks the “Stage to Hydroshare” or “Stage to Ikwai” menu items (Figure 7), a pop-up window appears showing all the metadata for the selection, along with a listing of all the files associated with this metadata. The user selects which files to include and then clicks the “Stage to Hydroshare” or “Stage to Ikwai” button (Figure 8).

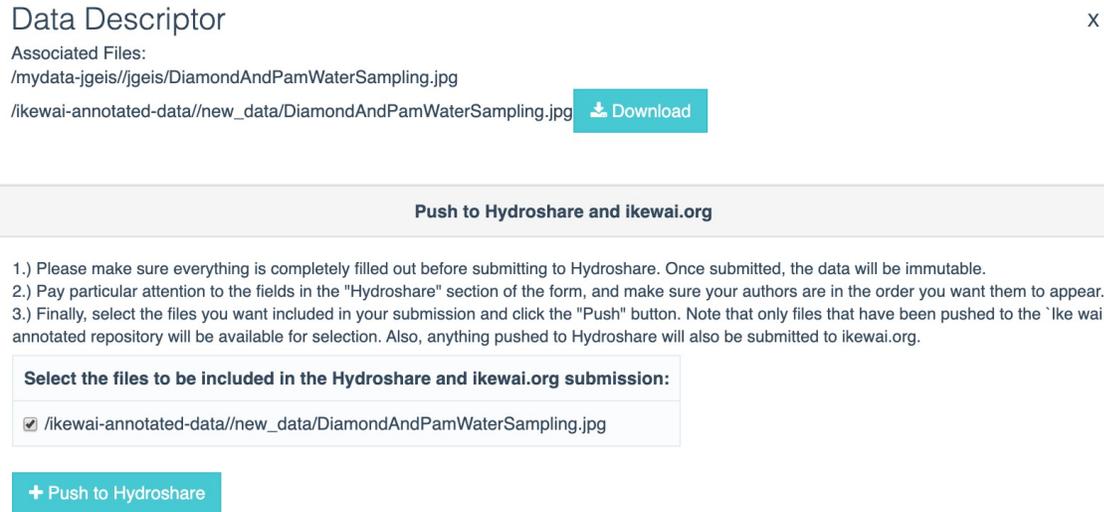


Fig. 8. Screenshot of the 'Ike Wai gateway's "Push to Hydroshare" page.

At this point in the dissemination workflow, the information has not actually been pushed. To ensure that the data is properly annotated, the submission goes through a “staging” process. The metadata for this submission gets flagged to indicate that it’s been staged. An administration page shows all staged submissions allowing a data manager to review the submission and either approve it or defer(reject) the submission (Figure 9).

Manage Staged Data Descriptors

Filter Current Local Search Results				
Title	Owner	Status	Summary	Actions
Diamond Tachera sampling rain water at UHH CC in August 2018	jgeis	Staged to Hydroshare	Diamond Tachera, Graduate Research Assistant, Dept. of Earth Sciences, SOEST, sampling rain water at the UH Hawai'i Community College Pāalamanui Campus in August 2018. Photo taken by Dr. Kiana Frank, Assistant Professor, PBRC	Actions ▾
Diamond and Sheree sampling rainwater on Mauna Kea	jgeis	Staged to Ikwai	A photo of Diamond and Sheree sampling rainwater on Mauna Kea.	Actions ▾
Notes on possible overlap between SGD and Hawaiian translations	jgeis	Staged to Ikwai Staged to Hydroshare	Informal notes taken during a short early exploration in May 2017 of potential spatial overlap between possible groundwater science	Actions ▾

Fig. 9. Screenshot of the 'Ike Wai gateway's "Staged Data Descriptors Management" page.

In the event a submission is missing necessary metadata, the data manager can choose to “reject” the submission by selecting the appropriate menu item from the “Actions” menu. This brings up a modal in which the data manager can enter text to inform the researcher what needs to be done before the submission will be approved (Figure 10).

Reason for rejection

The location and date information is missing.

Cancel
Submit

Fig. 10. Screenshot of the 'Ike Wai gateway's "Staged Data Descriptors Rejection" modal.

After the data manager has complete their message and submitted the form, a flag is set on the submission and the user is sent an email with the current state of their submission and includes the explanation of what is needed. The user can then make the necessary changes and resubmit, upon which the process starts again.

If the submission has sufficient metadata annotation, the data manager selects the appropriate “Push to Hydroshare” or “Push to Ikwai” menu item from the “Actions” menu. The submission’s flag is changed to mark that it has been pushed to the appropriate site. When the user next logs in, there will be a status tag letting them know the submission has been approved.

The process of “pushing” to the Ike Wai data gateway is straightforward. To make the files and metadata show up on Ike Wai data gateway setting a flag on the data descriptor to public is sufficient (Figure 2). The Ike Wai data gateway has access to the gateway’s storage systems and does a search for all records with the appropriate public flag.

2.4.1 Hydroshare. When designing the process for creating a Hydroshare resource, it was determined that having each user create their own account was unwieldy and would make it hard for others to find all the 'Ike Wai related resources. To solve this, an “ikewai” service account was made, under which all resources from the Gateway would be created. When a data manager approves a submission, the gateway would use the Hydroshare API to create a “resource” in the Hydroshare ikewai service account. To prevent issues with Hydroshare’s storage limit, we don’t copy the actual files/data over, instead we create permanent publicly accessible links to the files in the Gateway that allow the files to be accessed by anyone. These links are reference and along with all relevant metadata, are copied over to Hydroshare to create the resource. The new Hydroshare’s resource ID is stored in the Data Descriptor for later reference.

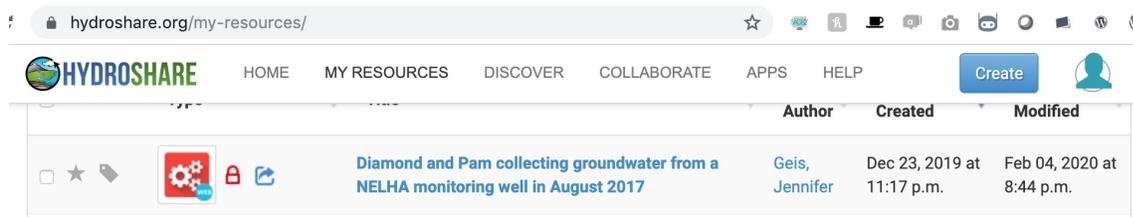


Fig. 11. Screenshot of Hydroshare’s “My Resources” page.

While using the Hydroshare API makes the transition of moving data from the private Gateway to the publicly accessible Hydroshare easier than a manual process, unfortunately, the data still will not be publicly accessible without one additional step. Files and metadata that comprise a Hydroshare resource are not visible to anyone outside of the Hydroshare account until the resource is “published” on Hydroshare. This action has the effect of creating a DOI, but publishing is not currently available via the Hydroshare API, only the web interface. To handle this, after accepting the Hydroshare submission, the data manager must log in to the ikewai service account on Hydroshare and publish the resource from there (Figure 11,12). After this, the data manager can access the Gateway and complete the workflow by marking the process complete, the gateway then pulls the assigned DOI based on the Hydroshare resource ID associated with the Data Descriptor.

2.4.2 Search and Discover. The 'Ike Wai data gateway supports two modes of search: a full-text search of the annotations and a spatial based search. Since the gateway metadata is stored in MongoDB, a full-text index has been enabled on the metadata fields allowing users to enter a phrase in the search interface that searches across all metadata fields for possible matches. The spatial search allows a geoJSON bounding area to query for matching metadata that has been spatial tagged. The Ike Wai data gateway user interface support both spatial and text search via an interactive map and search bar (Figure 13). This interface for data discovery support exploring the public metadata and download of file products that have been annotated and tagged as public.

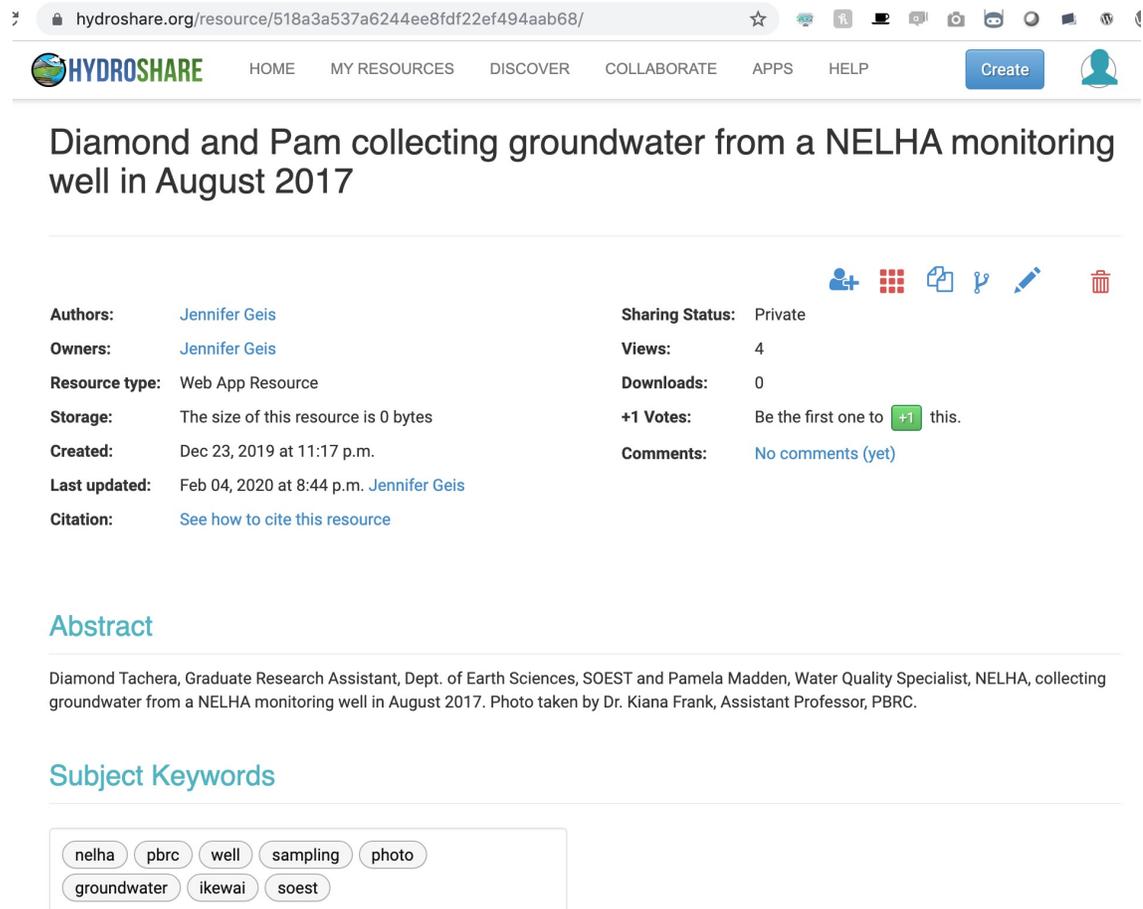


Fig. 12. Screenshot of Hydroshare’s edit resource page.

3 CONCLUSION

The ‘Ike Wai gateway, like all software, must continue to evolve to stay relevant to the changing needs of the the ‘Ike Wai user community. By leveraging the Tapis framework, UH has developed interface and workflows that provides effective annotation and dissemination features that encourage FAIR data practices. These gateway features support water sustainability research and aide agencies, water managers, researchers and the community in the state of Hawai‘i. Disseminating results for the ‘Ike Wai project through the ‘Ike Wai data gateway and Hydroshare make the research products accessible and reusable for stakeholders and the broader research community.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation Office of Advanced CyberInfrastructure[grant numbers 1931439 and 1931575], RII Track 1: ‘Ike Wai Securing Hawai‘i’s Water Future NSF OIA #1557349,

Browse 'Ike Wai Hawai'i Public Data Collection

Use the bounding tools on the left to create a bounding area to search

The screenshot shows a web-based search interface for the 'Ike Wai Hawai'i Public Data Collection'. On the left, there is a map of the Waimea area with a search box and a legend. A pop-up window titled 'Well' displays the following information:

- Name: Waimea C C
- ID: 8-6235-001
- Use: IRRGC
- Driller: Keauhou Kona Resort Co.
- Year Drilled: 1991
- Surveyor: null
- Casing Diameter: 14
- Depth: 1415
- Latitude: 20.04363
- Longitude: -155.596414

On the right, the search results are displayed in a table:

Results 2				
Waimea C				
	Name	Type	Lat,Long	Files
36155354701	8-6235-01 Waimea CC, HI	Water Quality Site	20.0433333, -155.5963889	
I	Waimea C C	Well	20.04363, -155.596414	Download 8-6235-001.pdf

At the bottom of the interface, there are logos for 'IKE WAI HAWAII', 'EPSCoR UNIVERSITY of HAWAII', and 'SSI'.

Fig. 13. Screenshot of Ike Wai data gateway search interface

Collaborative Research: SS2-SSI: The Agave Platform: An Open Science-As-A-Service Cloud Platform for Reproducible Science NSF ACI #145041

REFERENCES

- [1] Sean B. Cleveland et al. 2018. The 'Ike Wai Gateway- A Science Gateway For The Water Future of Hawai'i". Science Gateways Community Institute, Proceedings of Science Gateways 2018, Austin TX, USA September 2018.
- [2] E. Afgan et al. 2011. Galaxy: A Gateway to Tools in e-Science. Guide to e-Science. X. Yang et al., eds. *Springer London*. 2, 12 (2011), 48–53.
- [3] J. Towns et al. 2014. XSEDE: Accelerating Scientific Discovery. *Computing in Science & Engineering*. 5, 16 (2014), 48–53.
- [4] M.E. Pierce et al. 2015. Apache Airavata: design and directions of a science gateway framework. *Concurrency and Computation: Practice and Experience* 16, 7 (2015), 4282–4291.
- [5] R. Ananthkrishnan et al. 2015. Globus platform-as-a-service for collaborative science applications. *Concurrency and Computation: Practice and Experience* 2, 27 (2015), 290–305.
- [6] R. Dooley et al. 2002. Software-as-a-service: the iPlant foundation API. 5th IEEE Workshop on Many-Task Computing on Grids and Supercomputers.
- [7] S. Marru et al. 2011. Apache airavata: a framework for distributed applications and computational workflows. Proceedings of the 2011 ACM workshop on Gateway computing environments.
- [8] Lawrence K. Dahan M. Pierce M.E. Gesing, S. 2019. Science gateways: Sustainability via on-campus teams. *Future Generation Computer Systems* 94 (2019), 97–102.
- [9] Wilkins-Diehr N. Dahan M. Lawrence K. Zentner M. Pierce M. Hayden L. Gesing, S. and S. Marru. 2019. Science gateways: the long road to the birth of an institute. Proceedings of the 50th Hawaii International Conference on System Sciences.
- [10] et al. Goff, Stephen A. 2011. The iPlant Collaborative: Cyberinfrastructure for Plant Biology. *Frontiers in Plant Science* 2 (2011). <https://doi.org/10.3389/fpls.2011.00034>
- [11] S.J.K. Mason et al. 2014. A centralized tool for managing, archiving, and serving point-in-time data in ecological research laboratories. *Environmental Modeling & Software* 51, 59–69.
- [12] M. McLennan and R. Kennell. 2010. HUBzero: A Platform for Dissemination and Collaboration in Computational Science and Engineering. *Computing in Science & Engineering*. 2, 12 (2010), 48–53.
- [13] et al. Merchant, Nirav. 2016. The iPlant Collaborative: Cyberinfrastructure for Enabling Data to Discovery for the Life Sciences. *PLOS Biol* (2016). <https://doi.org/10.1371/journal.pbio.1002342>.
- [14] Joe Stubbs, Richard Cardone, Mike Packard, Anagha Jamthe, Smruti Padhy, Steve Terry, Julia Looney, Joseph Meiring, Steve Black, Maytal Dahan, Sean Cleveland, and Gwen Jacobs. 2020. Tapis: An API Platform for Reproducible, Distributed Computational Research. (2020). submitted.
- [15] D.G. Tarboton et al. 2009. Development of a community hydrologic information system. In *Proceedings of 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation, Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation*. 988–994.
- [16] R. Idaszak J. S. Horsburgh J. Heard D. Ames J. L. Goodall L. Band V. Merwade A. Couch J. Arrigo R. Hooper D. Valentine Tarboton, D. G. and D. Maidment. 2014. HydroShare: Advancing Collaboration through Hydrologic Data and Model Sharing. In *Proceedings of the 7th International Congress on Environmental Modelling and Software, San Diego, California, USA, International Environmental Modelling and Software Society (iEMSs)*.
- [17] N. Wilkins-Diehr and T.D. Crawford. 2018. The science gateways community institute and the molecular sciences software institute. *Computing in Science & Engineering* 5, 20 (2018), 26–38.
- [18] Mark D. Wilkinson, Michel Dumontier, Ijsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C. 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. 2012. WS-PGRADE/gUSE Generic DCI Gateway Framework for a Large Variety of User Communities. *Journal of Grid Computing*. 4, 10 (Nov. 2012), 601–630. <https://doi.org/10.1038/sdata.2016.18>
- [19] Mark D. Wilkinson, Michel Dumontier, Ijsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C. 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3, 1 (March 2016), 160018. <https://doi.org/10.1038/sdata.2016.18>