Received: 24 April 2020 | Revised: 6 November 2020 | Accepted: 20 November 2020
DOI: 10.1002/cpe.6130

WILEY

SPECIAL ISSUE PAPER

GHub: Building a glaciology gateway to unify a community

Jeanette M. Sperhac[1] | Kristin Poinar[2] | Renette Jones-Ivey[3] | Jason Briner[2] | Beata Csatho[2] | Sophie Nowicki[2] | Erika Simon[4] | Eric Larour[5] | Justin Quinn[5] | Abani Patra[6]

[1]Center for Computational Research, University at Buffalo, Buffalo, New York, USA
[2]Department of Geology, University at Buffalo, Buffalo, New York, USA
[3]Institute for Computational and Data Sciences, University at Buffalo, Buffalo, New York, USA
[4]Cryospheric Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA
[5]Sea Level and Ice Group, NASA Jet Propulsion Lab, Pasadena, California, USA
[6]Data Intensive Studies Center, Tufts University, Boston, Massachusetts, USA

Correspondence
Jeanette M. Sperhac, Center for Computational Research, University at Buffalo, 701 Ellicott St., Buffalo, NY, USA.
Email: jsperhac@buffalo.edu

Funding information
National Science Foundation, Grant/Award Numbers: 1837544, 2004826, 2004302

Abstract

There is no consensus on how quickly the earth's ice sheets are melting due to global warming, nor on the ramifications to sea level rise. Due to its potential effects on coastal populations and global economies, sea level rise is a grave concern, making ice melt rates an important area of study. The ice-sheet science community consists of two groups that perform related but distinct kinds of research: a data community, and a model building community. The data community characterizes past and current states of the ice sheets by assembling data from field and satellite observations. The modeling community forecasts the rate of ice-sheet decline with computational models validated against observations. Although observational data and models depend on one another, these two groups are not well integrated. Better coordination between data collection efforts and modeling efforts is imperative if we are to improve our understanding of ice sheet loss rates. We present a new science gateway, *GHub*, a collaboration space for ice sheet scientists. This web-accessible gateway will host datasets and modeling workflows, and provide access to codes that enable tool building by the ice sheet science community. Using GHub, we will collect and centralize existing datasets, creating data products that more completely catalog the ice sheets of Greenland and Antarctica. We will build workflows for model validation and uncertainty quantification, extending existing ice sheet models. Finally, we will host existing community codes, enabling scientists to build new tools utilizing them. With this new cyberinfrastructure, ice sheet scientists will gain integrated tools to quantify the rate and extent of sea level rise, benefitting human societies around the globe.

KEYWORDS
community codes, datasets, high-performance computing, ice-sheet science, science gateways, tool building

1 | INTRODUCTION

Global sea level rise is a serious concern and a topic of concentrated current study. Indeed, ongoing measurements of the mass of the Greenland ice sheet indicate that it contributes between 0.6 and 0.8 mm to global sea level *each year*.[1] However, recent measurements from the bedrock underlying the Greenland ice sheet suggest that modest warming in geologically recent times caused much greater ice loss and sea-level rise than is typically captured in ice sheet model simulations.[2] These data suggest that the current generation of ice sheet models substantially *underestimates* future

https://theghub.org

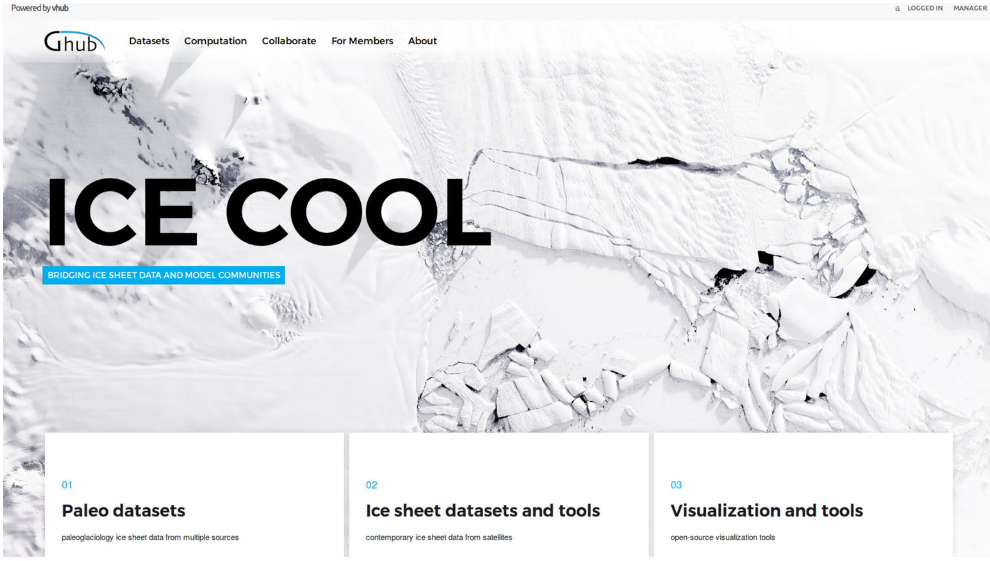

FIGURE 1 GHub user interface

sea-level rise, raising concerns about the potential effect of future warming on coastal populations, global economies, and national security. This scientific problem requires better coordination and improved collaboration from the involved scientific communities, as a recent NSF workshop concluded.[3] At present, the associated report determined, ice sheet science progress suffers due to siloed datasets stored in obscure locations, in nonstandard and varied data formats; isolated and special-case scientific codes; and limited access to computational resources. Tackling the problem of projecting ice sheet melt requires the concerted input of specialists in disciplines ranging from geology and climate and ice sheet modeling to glaciology, geophysics, sea level studies, and others.

A *science gateway* provides streamlined, user-friendly access to advanced scientific resources using just a web browser. Through its online interface, a gateway combines cyberinfrastructure components to provide researchers with access to diverse, discipline-specific resources that include computational resources, advanced software, collaboration tools, and data repositories.[4,5] With a gateway, isolated scientific codes developed by domain scientists can be colocated together with data, inputs, and documentation, becoming centralized computational tools that can benefit research and teaching by the entire community. A dedicated science gateway offers the ability to unify the ice sheet science communities, enabling collaborators to achieve a better understanding of these scientific challenges, and eventually, better solutions.

Our new science gateway, GHub,[6,7] will provide ready access to ice sheet data, software tools, community codes, collaborative tools and utilities, and educational materials via web browser; see Figure 1 for its user interface. GHub, which is presently under development, will enable new channels of communication and collaboration between data generators and modelers in the ice sheet scientific community. In so doing, it will improve capabilities for ice sheet modeling and help the community work to decrease uncertainty in sea level rise forecasting.

In the remainder of this article, we describe both our achievements in this project, and our plans. We introduce GHub's cyberinfrastructure (Section 2), then outline the GHub team's plans for supporting ice sheet science datasets (Section 3), tools (Section 4), and community codes (Section 5) on the new gateway. Finally, we highlight our accomplishments to date on these fronts (Section 6).

2 | PROVIDING COMMUNITY CYBERINFRASTRUCTURE

The ice-sheet community is in need of a central location in which they can coordinate the pursuit of new science. Science gateways excel at providing just that.[8] A science gateway typically provides a web-based user interface for a specific research community to access a set of curated applications tailored to its needs. Science gateways provide a straightforward way for users, especially those in disciplines that are not traditionally computational, to access and easily utilize high-performance and cloud computing.[9] To this end, the GHub gateway has been created by the University at Buffalo (UB) Center for Computational Research (CCR) and hosted at the San Diego Supercomputing Center. It offers a growing number of computational tools and datasets to support research and teaching by the worldwide ice sheet community. GHub is built on the HUBzero Platform for Scientific Collaboration, a proven open-source science gateway platform that enables users to launch advanced software tools and computations with a web browser, without having to download, compile, or install code on local systems.[10]

The ice sheet scientific community is small, numbering about 150 individuals, and yet is poorly unified as well as being widely dispersed around the world. The GHub gateway will enable scientists and members of the interested public to browse an attractive web interface (Figure 1) for computational tools, datasets, documentation, and resources that are available for ice sheet science. Registered users, once approved by the site administrator, are granted access to the resources deployed on the GHub gateway and can read documents, preview datasets, create tools, and run computations on high-performance computing (HPC) resources right from the web interface, according to their permissions.

2.1 | Cyberinfrastructure platform

The HUBzero platform is scalable and can serve a large user base. The flagship example of HUBzero's capabilities is NanoHUB (nanohub.org), which served 1,496,000 unique visitors in 2018. Of these, a core audience of 17,000 users ran computations with the site's 530 available tools. Similarly, in the VIDIA gateway run by CCR, the HUBzero platform has proven robust enough to support classroom instruction: in hands-on class sessions, 25 students log on simultaneously to run calculations in concurrent tool sessions.[11,12] Our experience suggests that the platform will be more than adequate for the needs of the comparatively small ice sheet science community.

Another example of a successful science gateway is provided by VHub.org. In 2010, CCR created this HUBzero-based science gateway for the volcano hazards field.[13,14] VHub, a virtual scientific community for sharing knowledge, tools, and educational materials, has become a true hub for volcano researchers, with 20,000–25,000 unique visits per month, substantial international participation, and approximately 10,000 members. VHub hosts complex, large-scale simulation tools developed and used by the volcanology community, such as a parallel adaptive mesh refinement mass flow simulator called Titan2D;[15,16] ensemble-based Bayesian approaches to uncertainty quantification; and probabilistic simulations of volcanic ash transport that are generated from multiple volcanic plume rise simulations.[17] The VHub gateway makes the complex interplay of data and simulation involved in these workflows seamlessly available to geophysical scientists. In addition, numerical models implemented on the VHub platform are used by government volcano observatories in nations such as Colombia and Indonesia for volcanic event disaster planning and preparedness.[18]

2.2 | Architecture and functionality

GHub's architecture consists of a database server and webserver; an execution host that runs software containers for computational tools; and middleware—software that coordinates the container sessions with user sessions. Computing cycles are provided by CCR's HPC cluster. In particular, containers running on GHub use submission software to queue large computing jobs with the cluster's resource manager. The cluster itself has direct access to GHub's data store, housed at CCR. GHub users can use their credentials to efficiently access the data store and computational results using Globus file transfer utilities.[19,20]

The gateway's webserver handles user accounts and interaction, including registering and subsequently authenticating users, controlling access to tools and other hosted resources, and providing collaborative features. On requesting an account, a user must indicate their understanding that the platform is to be used for academic purposes, then receive approval from a GHub administrator. Once approved, registered users are allocated a disk quota on the gateway, and can request more space as needed. Access to gateway tools and other resources are designated by the user that publishes the resource (see Section 2.3). The webserver also manages the collaborative features of GHub's HUBzero platform, including joint online calendars, forums, wikis, file sharing, project areas, and so forth.

Computational tools form the heart of the gateway. When a user runs a tool on GHub, a virtual container is started on the execution host. Each tool container has been configured to support specific computational needs, such as memory or disk space. Additional execution host servers may be deployed to scale up either the number of users supported, the resource footprint for tool sessions, or both. Finally, tools needing additional resources or parallel execution can submit jobs to CCR's HPC cluster. Configuring GHub to utilize different or additional computing resources is straightforward.

2.3 | Fostering scientific tool development

GHub enables its users to deploy their scientific codes as user-facing tools. GHub's ice scientist users can create and deploy computational tools on the gateway, using a containerized development environment styled like a Linux desktop. This full-featured environment offers multiple compilers along with debuggers and source code revision control. A development lifecycle guides users through a framework for publishing their tools on GHub, allowing them to control access to their tools. Thus, users may grant tool access to the gateway's whole user community, or limit it to individual groups or users. In addition, tools undergo vetting, testing, and review by domain scientists and GHub administrators prior to publication on the gateway. To supply GHub's tools with additional computational power, GHub tool developers can transparently submit jobs to CCR's HPC cluster. Furthermore, as the gateway's computational needs increase, GHub can easily be configured to utilize additional computing resources, such as public or private clouds or NSF's XSEDE.

Users may alternately use the gateway's Jupyter Notebooks environment to develop and share their computations, integrated with documentation, visualizations, and graphics, using Python and associated scientific software packages and libraries.[21,22] The standard scientific packages are provided on GHub (matplotlib, numpy, and so forth); integration with Anaconda streamlines the installation and management of additional packages.[23] For example, the open-source Hublib package, written in Python, allows tool developers to easily create dashboard-style front-ends for user interaction with Jupyter Notebooks.[24] The resulting notebooks and dashboards are both user-friendly and straightforwardly deployed as tools. The open-source Pegasus workflow framework is also supported on Ghub, enabling users to construct fault-tolerant scientific workflows from their codes.[25] Such computational workflows are a boon to disciplines such as ice sheet science, where multiple computational steps or resources may be pipelined together to achieve a research aim.[26] Used in either notebook or Linux desktop mode, the supported development process enables gateway users to craft their scientific codes into user-facing tools. For example, Section 4.1 describes the transformation of a scientific crevasse-detection code into a GUI-driven workflow tool hosted on the GHub gateway.

Importantly, codes installed on CCR's HPC cluster can become a focus of tool development efforts by the GHub gateway community. For example, ice sheet science community codes such as CmCt, ISSM, and SERAC, once installed on the cluster, become available for GHub's users to build computational tools against them. GHub's scientist users may build user interfaces to collect input parameters, relay the needed calls to the codes on the HPC resource, then gather and present the resulting output using plotting, charts, and the functionality of Jupyter Notebooks. Tool developers need not compile or install these codes, or secure time on HPC resources, in order to gain access to hosted codes and develop these tools. The tools built on such community codes are then published on GHub for others in the ice sheet science community to use. Refer to Section 5 for a discussion of the community codes that form a prominent part of this project.

3 | DATASETS

One prominent goal of the GHub project is to assemble and host ice sheet datasets. Ice sheet science data can be broadly divided into two types: (i) paleoclimatologic, and (ii) contemporary data. Paleoclimatologic (or *paleo*) ice sheet data, which is derived from a variety of geological methods, records *historical* ice sheet change, whereas contemporary ice sheet data is collected by modern satellite missions, and documents *current* ice sheet characteristics. Both types of data are crucial to shedding light on ice sheet sensitivity to climate change. However, there is currently no single virtual meeting place that serves scientists in the ice sheet data community. Though a number of active data centers currently serve the ice sheet, glacier, and polar science communities, each community's disparate needs send it to a specific data center. While ice sheet modeling groups collaborate during important model intercomparison projects, and observation-based research and paleo-based research pursue their own disciplinary efforts, no single location enables these communities to access, share, and exchange data products. For these reasons, the GHub team aims to host both contemporary and paleo datasets, making these ice-sheet datasets accessible to the modeling community in a single location. Offering both of these types of ice sheet data will help foster the appropriate use of datasets to validate models, enable the use of valid but esoteric datasets in modeling, and ensure the participation of modelers in the data collection process.

At the outset, GHub is envisioned as a read-only store of data from the perspective of most users. The GHub team's geologists and computer scientists are planning and preparing datasets and validating and uploading data, making it accessible to GHub's tools and registered users. Such users can employ Globus file transfer utilities to download data from CCR's GHub data store, as needed for their own analyses. The GHub data store is co-located at CCR with the academic HPC cluster used as the computational resource for the GHub gateway. This will keep data transfer to a minimum at computation time.

3.1 | Contemporary ice sheet data

Contemporary ice sheet data consists of airborne or satellite-based observations of the planet's ice sheets. These data capture measures such as ice sheet surface height or thickness; and ice flow speed. These observational datasets, compiled in major missions such as Landsat,[27,28] ICESat-2,[29] NASA Operation Ice Bridge,[30] and ESA Sentinel,[31] are collected and processed, then hosted for the public by data archival centers such as USGS-EarthExplorer,[32] the US National Snow and Ice Data Center (NSIDC),[33] and the European Space Agency's Copernicus Open Access Hub.[34] Such centers commonly provide web-based data viewers, download tools, and APIs. However, directory structures and file formats vary, meaning that dataset users must develop and maintain multiple data access procedures.

Ice sheet elevation change data provide sensitive indicators of ice sheet dynamics. These data can therefore assist scientists in validating models, but in order to do this, time-series datasets of these elevation changes are needed. These datasets need to be prepared, vetted, and hosted. Members of the GHub scientific team previously developed the SERAC (surface elevation reconstruction and change detection) method for characterizing ice sheet elevation changes.[35,36] SERAC is used to create these time-series datasets with increased spatiotemporal resolution and accuracy. The GHub scientific team will employ the SERAC approach to fuse existing datasets, generating elevation and thickness time series from a variety of observational missions spanning the last two decades. These datasets will then be hosted and documented on GHub for the community to utilize. The tools used to create them will additionally be integrated, expanded, and hosted on the gateway as well. Refer to Section 4 for detailed information on the initial toolset that is planned.

3.2 | Paleoclimatologic ice sheet data

Paleoclimatologic data, often referred to as "paleo" data, encode changes in the planetary climate, on the scale of the entire history of Earth. These data may be obtained using a wide variety of geological methods, and include data from rocks and sediments as well as from ice sheets, tree rings, microfossils, and other sources.

Reconstructing how ice sheets evolved to their present configuration is desirable for establishing ice sheet climate sensitivity and forecasting future behavior and sea-level contributions. Models based on paleo data can produce viable depictions of ice sheet evolution,[37] but their creation presents several obstacles. First, the paleo datasets are generated by geologists; ice sheet modelers may not always appreciate the limitations, and the appropriate use of, these datasets. Second, the geologists who produce these datasets are sometimes unaware of the data formats and resolutions that are most useful for modelers; the resulting task of appropriately formatting paleo data and preparing them for testing them against ice sheet models is currently left to the individual modeler. Third, modelers seeking paleo data may have difficulty accessing these data, since most are not stored in data archive centers, and may even be buried in the disciplinary literature.[38,39] Together, these obstacles inhibit the use of paleo ice-sheet data in the development and validation of predictive ice-sheet models.[3]

These issues of data discovery, access, and conversion create substantial roadblocks that impede the progress of ice-sheet modeling with paleo data, meaning that modeling efforts that use paleo data constraints are rare. With the GHub gateway, we hope to smooth the way for appropriate and collaborative use and easy access to these informative datasets for forecasting ice sheet melting. Dataset hosting, comprehensive documentation, and examples of use, all in a central location, will assist this diverse community in the creation and use of paleo datasets.

Three types of Greenland ice sheet paleo data are slated to be offered initially on GHub: relative sea level data from the Greenland periphery; ice margin mapping and dating from the literature; and Greenland global positioning system network (GNET) data.[40,41] All of these paleo data help to identify the extent of the ice sheet over time. For example, the Greenland GNET collects measurements from more than 60 GPS stations located around the edges of Greenland's ice sheet. These measurements indicate postglacial uplift rates, which allow ice-sheet scientists to infer past and present mass changes in the ice sheet.

3.3 | Intercomparison projects

Intercomparison projects in the ice sheet community are a key to understanding the current and future state of the ice sheets. These intercomparison activities are taking place in both the observation and modeling communities, yet remain disconnected from each other despite facing similar challenges. For example, the Ice sheet mass balance intercomparison exercise projects utilize multiple techniques to derive ice sheet mass balance estimates from remote sensing observations ranging from altimetry to gravimetry. The task of reconciling the resulting mass balance estimates requires concerted effort from the international scientific community.[42-44]

The latest ice sheet modeling intercomparison project (Ice Sheet Model Intercomparison Project 6, or ISMIP6)[45] employs thirty ice sheet models, deployed by groups from around the world, and all applying the same future climate scenarios to their ice-sheet models. It is the task of the intercomparison project to capture the uncertainty in sea level projections between models, using these simulations. These intercomparison efforts require substantial coordination by numerous national and international participants, including organization, protocol design, collection, manipulation, and analysis of large, complex datasets. Coordinating between participants, hosting and accessing data, and managing these sizeable projects requires collaboration tools and storage space, such as a gateway can provide.

The present ISMIP6 project is already making great use of the new GHub gateway. More than 10TB of fast, secure, backed-up disk space has been allocated to the project at UB's CCR, where it is read-accessible to registered users via Globus data transfer tools, using CCR's new capability as a Globus endpoint.[19,20] Designated project data administrators have write access and final approval over these files. They have uploaded fully vetted ISMIP6 model output to the project storage, and all GHub and ISMIP6 members have read-only access to these data. Refer to Appendix A for details about data format and directory structure.

More than 60 members of the worldwide ice sheet modeling community have registered on GHub to use its collaborative tools, verify models, and access datasets. Furthermore, we are engaged in creating online tools and utilities that can assist modelers in verifying, checking, converting, and annotating their models as a part of this ambitious project. The GHub team is looking ahead to the ISMIP7 intercomparison project and considering various computational approaches to streamlining data management as well as project design and management for this distributed global undertaking. (See Section 4.2 for an outline of the tools proposed for the Ice Science Modeling Intercomparison Projects (ISMIP).)

4 | TOOLS

The GHub gateway supports creation of tools and workflows with its tool-building features, Pegasus[25] workflows, and scientific libraries for Python and other languages, as discussed in Section 2.3. Ice sheet scientists will use these features to develop software tools that make use of the ice sheet dataset products described in Section 3. The ice sheet scientific community can then access these software tools for research, teaching, and scientific projects such as ISMIP.

During the initial phase of the GHub project, the project team has used the HUBzero platform's Jupyter Notebooks capabilities to develop a series of pilot tools based on Python scripts and common workflows. Refer to Table 1 for a list of these tools, with status and tool descriptions. Two of the tools, *crevasseoib* and *netcdfregrid*, are production-quality and are suitable for researcher use; the remaining pilot tools are proofs of concept that allowed us to demonstrate the ease of developing Jupyter Notebook-based tools and outline our plans for tool pipelining to support the ISMIP6 project effort. Refer to Sections 4.1 and 4.2.2 for further information about *crevasseoib* and *netcdfregrid* and Section 4.2 for more about the ISMIP6 tools.

4.1 Crevasse detection workflow

During the GHub pilot, we developed a workflow tool that automates the detection of anomalously large crevasses in the Greenland ice sheet. This tool is particularly applicable to Southeast Greenland, where vast subsurface liquid water stores known as firn aquifers form.[46] These aquifers originate when high winter snowfall rates cover summer meltwater, insulating it from cold air temperatures that would otherwise refreeze the water. Though invisible from the surface, firn aquifers drain through crevasses to the glacier bed, influencing the flow of the ice above.[47] The resultant large crevasses are visible from the surface. Drainage of the firn aquifers changes the organization of the subglacial water system in ways that can stabilize or destabilize the flow of the ice above.[48] The timing, duration, and magnitude of these drainage events are currently unknown, and as a result, these processes are absent from current ice sheet models.[48,49]

Computer scientists and geologists used the GHub platform in collaboration to develop a crevasse detection workflow that provides a means to infer the drainage patterns of the firn aquifers around Greenland in space and time. The workflow consists of multiple, formerly user-intensive processes that are pipelined together on GHub (see Figure 2). The heart of the calculation is a script previously developed by a domain scientist to detect elevation anomalies in NASA IceBridge Airborne Topographic Mapper (ATM) datasets.[30,50,51] The tool's Jupyter Notebook front-end prompts users to enter the year range, crevasse detection tolerances, and latitude/longitude bounds of interest into a GUI created using Hublib components.[24] The workflow then uses Pegasus, Slurm,[52] and Python, and allocates compute nodes on the CCR's HPC cluster, from which it accesses, processes, and grids the appropriate IceBridge ATM datasets from the NSIDC.[53] It then runs the crevasse-identifying analysis code developed by Petty et al.[50] Finally, the workflow returns the results for the user to browse, view, and download. Results include analysis, plots and text files of crevasse size, shape, depth, and locations. Refer to Section 2.3 for a description of GHub's tool development and deployment infrastructure used to develop this tool.

Previously, substantial user interaction was required to complete each of the steps in this computation. The user was required to collect the appropriate datasets according to year and location parameters, secure time on an HPC resource, and submit the calculation after transferring the data, parameters, and scripts to the cluster. Then, once the job ran on the cluster, the user had to download the results, plot them, and finally analyze the results. To run this new workflow, however, the user need only interact with a Jupyter Notebook running in a web browser. The workflow pushes both the management of the job and the data work into the background, allowing the scientist to focus on their analysis. The workflow can be readily shared along with documentation, datasets, figures, peer reviewed papers, and conference presentations, for reproducible and shared research outcomes that accelerate community science.

The crevasse tool has already been used in a student research project evaluating spatial correlations between crevasses and the firn aquifer in the Puisortoq region, near the southern tip of Greenland.[54] The project detected wide crevasses directly downstream from the firn aquifer boundary along 100% of the ATM flight lines in the study region. Inside the aquifer bounds, the workflow detected only narrower crevasses and some false positives, including undulating surface terrain. The latter may point to an improvement that can be made in the detection algorithm. Researcher

TABLE 1 Initial Jupyter Notebooks based tools on the GHub science gateway, with their status

Tool name (Paper Section)	Short name	Description	Status
Workflow for crevasse detection (Section 4.1)	*crevasseoib*	Detects crevasses with specific criteria in satellite datasets	Production
netCDF File Regrid Tool (Section 4.2.2)	*netcdfregrid*	Batch regrids multiple netCDF files	Testing
ISMIP6 Model Comparison Tool (Section 4.2.1)	*gisplot2*	Compares a selected field between multiple ice sheet models	Pilot
CmCt Histogram Tool (Section 4.2.1)	*cmcthistplot*	Plots selected CmCt comparison results	Pilot
ISMIP6 Model Scalar Plotter (Section 4.2.1)	*scalarplot*	Plots selected scalar from ISMIP6 model	Pilot
CF-Processing Tool (Section 4.2.1)	*cfcompliance*	Processes ISMIP6 netCDF files	Pilot
CF-Regridding Tool (Section 4.2.1)	*cfregrid*	Regrids single netCDF file	Pilot

Note: The section of the paper in which each tool is discussed is indicated.

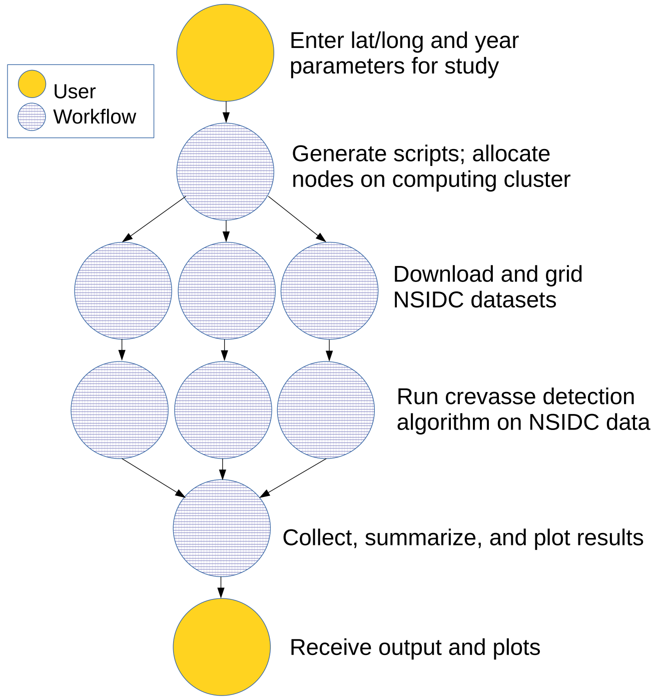

FIGURE 2 GHub crevasse detection workflow schematic. Yellow, user tasks; blue hatching, tasks automated in software, illustrates how data and high-performance computing resource management are handled by the workflow. Previously, user-scientists performed all steps of the workflow manually.

feedback about output formats was used to improve the tool, converting some output into comma-separated value format for easier incorporation into downstream analyses.

4.2 | Tools for ISMIP projects

The ISMIP compare ice sheet model simulations that have all run the same suite of experiments. These projects involve the cooperation and collaboration of a globally distributed team of ice sheet modeling experts who must calculate, prepare, vet, verify, grid, and upload data and models. The entire effort is overseen by a steering committee charged with tasks from planning and management to scientific protocols and data quality control. Each iteration of this project involves an international team, encompasses terabytes of data, and spans 5+ years, and must confront these issues:

- Data collection, validation, gridding, data analysis, and hosting;
- Fostering broad scientific access to project output; and
- Management: organizing, guiding, and planning the project.

The ISMIP6 steering committee is interested in using GHub for the planning and execution of its follow-on effort, ISMIP7. Two substantial contributions to the project were made during the GHub pilot: First, using UB CCR's storage for uploading ISMIP6 ice sheet model simulations and downloading the climate forcing datasets (see Section 3.3); Second, GHub's pilot validation tools that simplify quality control checks on the data submissions (see Section 4.2.1). Looking forward, project management tasks for ISMIP can be substantially simplified using the centralized online location and toolset provided by the GHub gateway (see Section 4.2.3).

4.2.1 | Data validation and verification

A critical component of the ISMIP projects is the painstaking verification of the data contributed to the project. Previously, the responsibility for data verification lay with the ISMIP project steering committee; GHub will help them by hosting standard tools that enable data contributors themselves to perform the needed actions. During the GHub pilot, the team developed several trial tools designed to assist ISMIP6 contributors in the verification and validation of their data; see Figure 3 for a visualization of an ice sheet model using one such tool, the pilot ISMIP6 ice sheet model comparison tool. Additional Jupyter Notebook tools were developed to assist in tasks such as regridding ice sheet model submissions (CF-Regridding Tool) and performing calculations for data quality control and standard compliance (Histogram Tool, ISMIP6 Ice Sheet Model Scalar Variable Plotter, and CF-Processing Tool).[55,56]

These tools are based upon well-vetted Python scripts developed by ice sheet scientists and used regularly in their work. As with the workflow described in Section 4.1, the scripts, once deployed as tools, become available to the scientific community in a standard location. At present, these

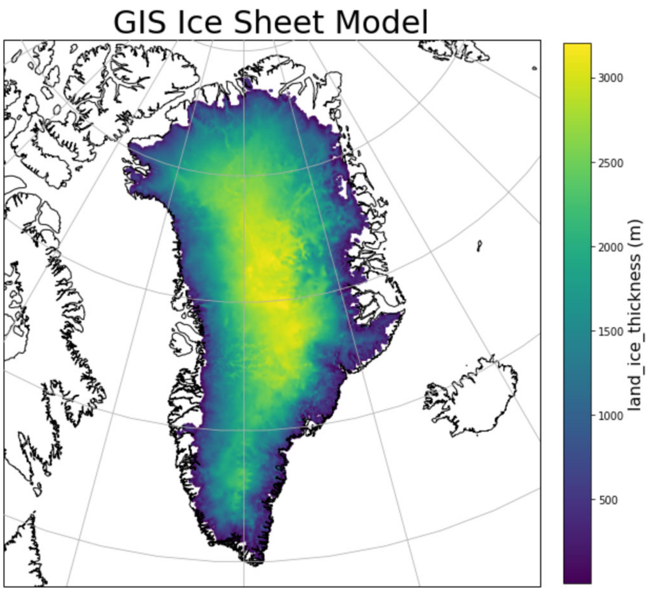

FIGURE 3 ISMIP6 project ice sheet model of Greenland, visualized using the GHub pilot gisplot2 tool

tools are in pilot form, developed to verify that the functionality and user interface needed by the community is possible. The goal is for such quality control tools to allow contributors to verify that their submissions have the correct magnitude, trends, spatial patterns and units. Additional tools will allow modelers to check that their submission is climate and forecast compliant. Participating ISMIP7 modelers will be asked to check their own submissions, correcting and rewriting any noncompliant ones, prior to submission. Eventually, a GHub-based software pipeline will provide data verification, implementation of corrections, rechecking, and submission stages, helping both contributors and the steering committee to ensure that the ISMIP data has the highest quality.

4.2.2 | Data exploration and availability

Some 30 different ice sheet models have contributed to the current ISMIP6 endeavor (see Section 3.3), and after the effort of processing and analyzing this substantial dataset, the results should ideally be made available for community use. Unfortunately, ice sheet science does not yet enjoy access to community tools that allow exploration and analysis of large model datasets. The GHub gateway can help address this problem, by hosting processed datasets and providing new tools designed for browsing, downloading, converting, and analyzing these data. Tools such as the batch regridding tool, piloted on GHub and in testing now, will help. Such tools will open up this valuable data resource to the rest of the ice sheet community, as well as other communities interested in sea level science.

The *netcdfregrid* tool was created on GHub to assist modelers who use ISMIP6 data products as input to their models. This tool converts multiple existing ISMIP6 data files to the user's preferred kilometer grid size. At the core of the tool is a first-order conservative remapping performed by calls to the CDO Python library's *remapcon* operator. This remapping task, which preserves integrals over the data between the source and target grids, is the standard regridding method used in the ISMIP6 project. The calls to this transformation are themselves simple, but require multiple files: the netCDF data file for remapping, as well as input and output grids providing details about each grid point in the data file. The conversion process can be resource-intensive; furthermore, modelers are likely to regrid numerous data files. The batch regridding tool streamlines the modeler's workflow by managing the submission of potentially hundreds of jobs to the HPC cluster. Importantly, rather than transferring data and grid files for the regridding operation, the computation is performed on the academic HPC cluster colocated with the data storage for the ISMIP6 project. To operate the tool, the user selects files or entire directories using a Jupyter Notebook GUI; the tool then constructs the job script and submits the regridding command to the cluster. Once the computations are complete, the tool automatically emails the user. The user can then download the regridded ISMIP6 data from CCR, using a convenient Globus endpoint, and use the data for their own analyses. This tool is in acceptance testing on GHub, and has the potential to substantially streamline data preprocessing for modelers.

To this end, we are planning a series of GHub tools that enable seamless intercomparison between models and observations, and between models and other models. The building block for these analysis and plotting tools will be Jupyter Notebooks. We will augment existing codes with analysis and plotting tools used by the climate community to create tools that can manipulate and postprocess the ISMIP datasets.[57,58] For example, scientists have requested access to ISMIP6 results by integrated basin scale ice mass change instead of the available ice sheet wide numbers. The basin scale change in ice mass information can be obtained from postprocessing the ice sheet 2D changes in ice thickness, and a tool in GHub would allow any user to easily define their regions of interest and carry out the postprocessing themselves. Other tools will enable dataset plotting and evaluation, multimodel analysis, or intercomparison with observational datasets. These tools will ensure the ISMIP data enjoys heavy use in the community. The flexibility of Jupyter Notebooks allows users to modify and build on such existing codes in order to make these datasets available to interested scientists.

4.2.3 | Project management

An interdisciplinary community effort like ISMIP6 requires coordination between more than 50 national and international members over the full life cycle of the project: the first ISMIP6 workshop was held in July 2014, and ISMIP6 is now at the stage where model projections are uploaded and analyzed. Many steps are required, including design and communication of protocol, evaluation of polar climate in models to be used,[59] preparation of climate dataset for use by ice sheet models, development of ice-ocean parameterizations,[60,61] dataset testing, recording and checking submissions, and soliciting and receiving feedback from scientists. During the first 5 years of the project, ISMIP6 management employed a variety of project management approaches, including emails, shared spreadsheets, and wikis, but this ad hoc strategy frequently resulted in miscommunication or duplication of efforts. For these reasons, the ISMIP6 steering committee plans to utilize the collaborative project management and processing capabilities of GHub, as outlined in Section 2. GHub's parent infrastructure offers unified project and upload spaces, calendars, wikis, and other utilities that will help coordinate and plan the follow-on effort, ISMIP7. The needed features are now in place on the GHub site.

The proposed ISMIP tools and project support in GHub will enable the ice sheet and polar modeling community to more easily manage the projects that produce these novel datasets. These datasets will then be hosted for the community to explore, visualize, and use in their ice sheet scientific endeavors. This will ensure the widespread use of these ISMIP datasets, which hold great promise for reducing the uncertainty in sea level projections, and advancing our understanding of the current and future state of the polar regions.

5 | COMMUNITY CODES

Computational tools are essential for modelers, ice sheet data scientists, and geologists to understand, process, assess, and validate the numerous types of ice sheet datasets that GHub is collecting. For this reason, developing and hosting community codes and computational tools based on them, is is an essential part of the GHub gateway's goals. The gateway will host selected existing community codes, with documentation, examples, and scholarly background, in a location where ice sheet modelers can easily upload their model results for comparison and validation, and access the codes for research and for further extension. The scientific community can then use these software tools, and the community codes hosted on the gateway, to investigate new metrics for model-data comparison and for model assessment, and to enable quantification of model uncertainties. Several codes have been identified for inclusion in the early stages of this project.

5.1 | Cryosphere model comparison tool

The cryosphere model comparison tool (CmCt) allows ice-sheet modelers to compare their own models with existing observational data from Greenland or Antarctica.[62,63] The tool, which uses an HPC cluster for its computations and is accessible via a web-based front end, is one of several community codes that has been identified for inclusion in the GHub gateway. It accepts a modeled ice sheet surface elevation, and performs an actual point by point comparison between the model and the observed surface elevation data from satellite datasets such as ICESat.[64] CmCt also allows comparison with gravimetry data from the gravity recovery and climate experiment (GRACE) mission using GRACE mascon and spherical harmonics solutions, as well as laser (ICESat-GLAS) and radar (ERS1, ERS2, Envisat) altimetry data.[65] The CmCt tool produces an output package that contains statistical information, along with model and observation differences. Notably, this tool carries out its comparison on the actual measurement that was made on a user-selected year.

The scientific team will expand CmCt significantly, open-source the resulting source code, and install it on CCR's HPC resource for use by the GHub gateway. In so doing, we will enable CmCt to directly compare ice-sheet models to observations of past ice-sheet change, and provide model performance metrics based on data-model misfit. To simplify data preparation for modelers, we will adapt CmCt to allow submission on a model's actual mesh, instead of requiring standard ice sheet gridding. We will streamline the dataset upload process, and enable intercomparison with user-supplied datasets, thus expanding the tool's comparison capabilities to any spatial dataset, regardless of the source. Furthermore, we will build in SERAC processing capabilities, letting the user define the spatio-temporal bounds for the comparisons. The enhanced CmCt source code will be installed and hosted on CCR's HPC cluster, supplemented on GHub by a Jupyter Notebook based user interface that will enable the direct plotting and analysis of CmCt results. (Several pilot tools in Table 1 are proofs of concept for analysis features.) The resulting CmCt tool will serve both modelers, who will run the enhanced online tool from GHub, and tool builders, who will develop new tools that call the underlying source code (see Section 2.3). These enhancements to the CmCt tool will equip ice sheet scientists to better link ice sheet forcing to the resulting response, thus enabling improved forecasts for melt and subsequent sea level rise.

5.2 | Ice sheet system model and virtual earth simulation laboratory

Ice sheet system model (ISSM) is a sophisticated ice sheet modeling package that requires installation on advanced computing resources such as a HPC cluster in order to harness its full capabilities. The virtual earth simulation laboratory (VESL)[66] is a website that allows users to explore visualizations generated by the complex ISSM model, using real earth science datasets. Both pieces of software were developed and open-sourced by NASA Jet Propulsion Laboratory (JPL).[67,68] Refer to Figure 4 for a visualization of an ISSM simulation of Greenland ice sheet velocities created using VESL.

VESL, hosted by JPL, provides a simplified, web-enabled interface to the ISSM package that exposes a subset of its controls with guided simulations. It utilizes Amazon Web Services to provide the needed HPC cycles for the package's compute engine. VESL is currently limited to a few types of configurations and a limited set of simulation options. A model of the Columbia glacier is one example of the web-enabled VESL tools that is relevant to the ice sheet question.[69] This and other VESL simulations expose ISSM model parameters such as controls for surface mass balance. This is done using real ice sheet data and live computation. Such simulations, which demonstrate some of the capabilities of the powerful underlying software package, are suitable for both outreach and education, and for use by scientists with specific needs.

With the GHub gateway, we aim to expand access to the powerful ISSM software package. The ISSM package is now installed on the CCR HPC cluster, making it accessible to GHub's community of tool builders and command line simulation users (see Section 2.3). Hosting the ISSM package on our HPC resource assists scientists by providing straightforward access to the capabilities of this ice sheet modeling code, including such functionality as stress balance, mass transport, static inversions, and mesh generation, without introducing concerns about obtaining access to suitable hardware and installing and compiling ISSM's source code.

As part of the GHub project, we will also extend VESL in several ways. We plan to extend support for additional glacier and ice sheet configurations in the tool. The visualizations currently available to users on VESL will be expanded to reflect new ice sheet simulations derived by scientists using the full ISSM package. To encourage more scientists to develop their own visualizations on GHub, we will develop an online, tutorial-driven modeling tool that will allow users to upload or select existing model or domain files, add and edit fields, adjust parameters, and then run solutions on various fields using the underlying ISSM software. The resulting web-based ISSM tool will enable straightforward interaction with the VESL's core engine within GHub itself. This new tool will better support outreach, education, and access to advanced ice sheet modeling by scientists with limited computing resources. With these extensions to VESL, users will have new visualizations to explore, and will gain confidence in using the underlying software to model their data.

5.3 | Surface elevation reconstruction and change detection

The GHub scientific team's prior work in ice sheet surface elevation reconstruction led to the development of the SERAC detection software tools[35] (see Section 3.1). SERAC enables users to significantly increase spatiotemporal resolution and accuracy when characterizing ice sheet elevation changes.[36] Several features offered by SERAC enable users to combine datasets from different sources and resolutions. Ice sheet altimetry data and digital elevation models (DEMs) have different spatiotemporal sampling and accuracy. When these data and models are combined, the

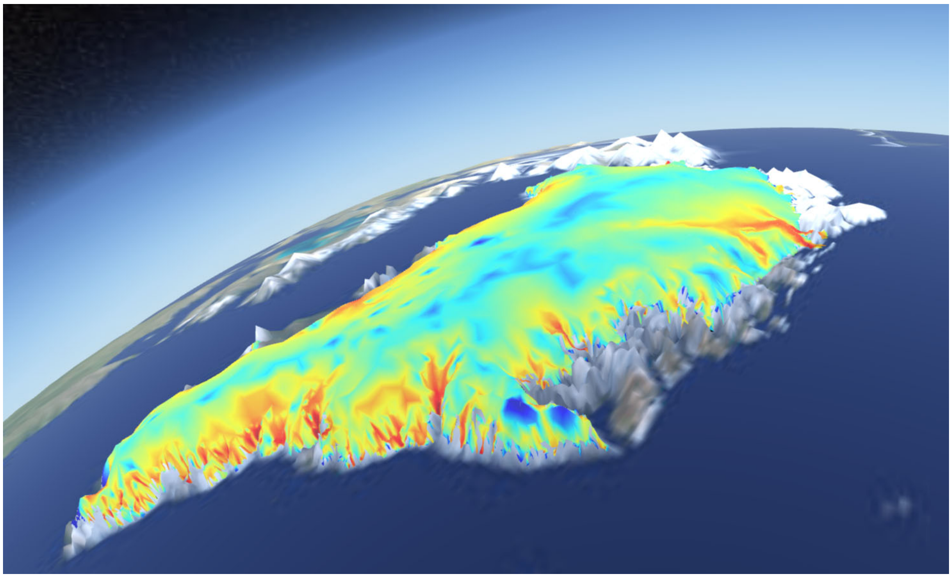

FIGURE 4 Visualization of Greenland ice sheet velocity in meters/year, for a specified basal friction value. Simulation was run on VESL, NASA JPL's online ISSM tool. Colors range from dark blue (0 m/year) to dark red (3500 m/year)

resolution typically suffers. Uniquely, SERAC allows users to preserve the original data resolution of different datasets, when combining sources with different resolutions. Moreover, SERAC also enables users to correct less accurate data, for example, DEMs derived from stereo satellite imagery.[36] However, the SERAC package includes a variety of software tools that are have not been tightly integrated, which can result in cumbersome and time-consuming processing.

To address this shortcoming, we will implement a complete SERAC workflow in GHub and CmCt. Several useful features will be incorporated into the new SERAC workflow, which will be open-sourced and installed on CCR's HPC cluster for use by GHub (See Section 2.3). The new SERAC workflow will support both altimetry and DEM data, as well as correction of user-defined datasets. Tools to interpolate the irregularly distributed observations into user-defined meshes or grids, to estimate ice elevation and thickness change, and to perform spatial interpolation will be incorporated into the workflow. Finally, elevation change time series will be combined with models enabling estimation of mass change time series.[70,71] By deriving time series of elevation and mass changes with error estimates to decadal changes of an entire ice sheet, the resulting tool will provide much-needed flexibility for intercomparisons, both between observations and between models and observations.

In addition to the enhancements to the SERAC tool, new data products, such as elevation change grids and mass change time series, will be created using SERAC and hosted on the GHub gateway. (Refer to Section 3.1.) Using the gateway, we will deploy time series data derived from NASA's laser altimetry observations, updating the datasets as new measurements are released.

6 | PROGRESS TO DATE

The GHub gateway is designed to facilitate new channels of communication and collaboration between data generators and modelers in the ice sheet scientific community. Our cyberinfrastructure framework will provide scientists with ready access to ice sheet data, software tools and workflows, high-quality community codes, HPC, collaborative utilities, and educational materials. The collaboration's goals promise to unify several scientific communities and make new data sources and computational tools available to researchers, educators, and students. A number of these goals have already been achieved.

During the year-long initial pilot phase of the GHub gateway, the project team has:

1. created the cyberinfrastructure framework for the GHub gateway (Section 2); including the tool container system, collaborative features, Jupyter Notebooks, and Workspace tools;
2. registered more than 60 gateway users from the ice sheet scientific community, many from the ISMIP6 project;
3. housed the 10TB ISMIP6 data release on accessible, backed-up media at UB CCR (Section 3.3);
4. piloted use of Globus tools on CCR resources, for ice sheet community members and gateway computational tools to access datasets such as ISMIP6;
5. piloted seven computational tools, including the crevasse workflow (Section 4.1), a batch regridding tool for ISMIP6 (Section 4.2.2), and several standalone pilot tools for ISMIP6 data validation and CmCt proofs of concept (Section 4.2.1 and Figure 3);
6. facilitated access to existing community codes such as ISSM, with installations on the CCR HPC cluster (Section 5.2); and
7. enabled collaborative features in a user-friendly web interface with an appealing custom design (Figure 1 and Sections 2 and 4.2.3)

7 | CONCLUSION

The GHub science gateway will present ice sheet science community users with datasets, tools and codes, and computing access in a single, unified location. Benefits GHub will confer to the ice sheet community include:

- A central location for domain datasets, documentation, examples, and underlying scholarship;
- Streamlined access to shared HPC resources and standardized workflows;
- A robust platform for teaching, workshops, and outreach;
- Broadened access and faster turnaround to scientific results; and
- Sharing and reuse of modeling experiments and data analysis tools.

Progress in ice sheet science suffers because of organizational and subdiscipline divides; barriers to access and accessibility of multiple types of data; obstacles to computation; and communication hurdles. These concerns can be mitigated by introducing a central location for ice sheet scientists to collaborate, use computational tools, and search for, convert, and compare data. The GHub gateway is poised to tackle these concerns. The ice sheet scientific community will benefit from the coordination and collaboration tools to assist in planning and executing large intercomparison projects, synthesis of paleo data, and development of new models. Science gateways provide community cyberinfrastructure with access to shared data, software and computational tools, computing resources, and collaborative features.[8] GHub, our new science gateway, will provide the galvanizing force to unify and strengthen the ice-sheet science community, providing a versatile new tool to improve coordination toward the question of ice-sheet melting rates and consequent sea-level rise estimates.

ACKNOWLEDGMENTS

Computing support for this project was provided by the Center for Computational Research at the University at Buffalo.[72] We thank the entire GHub scientific team, including Denis Felikson, William Lipscomb, Isabel Nias, Alek Petty, Joerg Schaefer, and CCR summer intern Elliot Snitzer. The work described here is supported by U.S. National Science Foundation (NSF) grants, “EAGER: Exploring a community driven data-model framework for testing the stability of the Greenland ice sheet,” NSF 1837544 and “Ghub as a Community-Driven Data-Model Framework for Ice-Sheet Science,” NSF 2004826 and NSF 2004302.

ORCID

Jeanette M. Sperhac https://orcid.org/0000-0003-0436-9318

REFERENCES

1. Velicogna I, Mohajerani Y, Geruo A, et al. Continuity of ice sheet mass loss in Greenland and Antarctica from the GRACE and GRACE follow-on missions. *Geophys Res Lett*. 2020. https://doi.org/10.1029/2020GL087291.47(8):e2020GL087291
2. Schaefer JM, Finkel RC, Balco G, et al. Greenland was nearly ice-free for extended periods during the Pleistocene. *Nature*. 2016;540(7632):252-255.
3. Briner J.P., Alley R.B., Bender M.L., Csatho B., Poinar K., Schaefer J.M. ‘How stable is the Greenland ice sheet?’. : NSF Workshop on Greenland Ice Sheet stabilityUniversity at Buffalo; 2017. This work was supported by National Science Foundation grant ‘The Stability of the Greenland Ice Sheet,’ award No. OPP-1741833. http://hdl.handle.net/10477/82467.
4. Science Gateways Community Institute (SGCI) https://sgci.org. Accessed March 12, 2020.
5. Wikipedia https://en.wikipedia.org/wiki/Science_gateway. Accessed March 12, 2020.
6. GHub project team GHub gateway; 2019. https://vhub.org/groups/ghub/. Accessed March 12, 2020.
7. Sperhac JM, Poinar K, Jones-Ivey R, et al. GHub: building a glaciology gateway to unify a community (talk) Gateways Community Institute’s Gateways 2019 Conference; 2019. https://doi.org/10.17605/OSF.IO/JGHBZScience
8. Wilkins-Diehr N, Zentner M, Pierce M, et al. The science gateways community institute at two years. *PEARC ’18XSEDEAssociation for Computing Machinery*. New York, NY: Springer; 2018.
9. Wilkins-Diehr N. Measuring success: how science gateways define impact; 2019. https://osf.io/tkzuy.
10. Mclennan M, Kennell R. HUBzero: a platform for dissemination and collaboration in computational science and engineering. *Comput Sci Eng*. 2010;12(2):48-52.
11. SUNY UB CCR. VIDIA Gateway; 2002. https://vidia.ccr.buffalo.edu/. Accessed September 7, 2020.
12. Sperhac JM, Gallo SM. VIDIA: a HUBzero gateway for data analytics education. *Futur Gener Comput Syst*. 2019;94:833-840. https://doi.org/10.1016/j.future.2018.02.004.
13. Volcano Hub (VHub) https://vhub.org/. Accessed March 20, 2020.
14. Palma JL, Courtland L, Charbonnier S, Tortini R, Valentine GA. VHub: a knowledge management system to facilitate online collaborative volcano modeling and research. *J Appl Volcanol*. 2014;3(2):1–16. https://doi.org/10.1186/2191-5040-3-2.
15. Patra AK, Bauer AC, Nichita CC, et al. Parallel adaptive numerical simulation of dry avalanches over natural terrain. *J Volcanol Geotherm Res*. 2005;139(1):1-21. https://doi.org/10.1016/j.jvolgeores.2004.06.014.
16. Simakov NA, Jones-Ivey RL, Akhavan-Safaei A, Aghakhani H, Jones MD, Patra AK. Modernizing Titan2D, a parallel AMR geophysical flow code to support multiple rheologies and extendability. *High Performance Computing*. Lecture Notes in Computer Science Vol 11887. Cham, Switzerland: Springer International Publishing; 2019:101–112. https://doi.org/10.1007/978-3-030-34356-9_10.
17. Bursik M, Jones M, Carn S, et al. Estimation and propagation of volcanic source parameter uncertainty in an ash transport and dispersal model: application to the Eyjafjallajokull plume of 14–16 April 2010. *Bull Volcanol*. 2012;74(10):2321-2338. https://doi.org/10.1007/s00445-012-0665-2.
18. Comisión Científica Nacional, Bogotá, Colombia . Servicio Geológico Colombiano (Colombian Geological Survey); 2020. https://www.sgc.gov.co. Accessed September 29, 2020.
19. Foster I. Globus online: accelerating and democratizing science through cloud-based services. *IEEE Internet Comput*. 2011;15(3):70-73. https://doi.org/10.1109/MIC.2011.64.
20. Allen B, Bresnahan J, Childers L, et al. Software as a service for data scientists. *Commun ACM*. 2012;55(2):81-88. https://doi.org/10.1145/2076450.2076468.
21. Project Jupyter Jupyter Lab https://jupyter.readthedocs.io/en/latest/. Accessed September 7, 2020.
22. Pérez F, Granger BE. IPython: a system for interactive scientific computing. *Comput Sci Eng*. 2007;9(3):21-29. https://doi.org/10.1109/MCSE.2007.53.
23. Anaconda Anaconda Software Distribution Computer software Vers. 2-2.4.0. Anaconda, November. 2016. https://anaconda.com.
24. HUBzero Hublib UI https://hubzero.github.io/hublib/. Accessed September 7, 2020.
25. Deelman E, Vahi K, Juve G, et al. Pegasus: a workflow management system for science automation. *Futur Gener Comput Syst*. 2015;46:17-35. Funding Acknowledgements: NSF ACI SDCI 0722019, NSF ACI SI2-SSI 1148515 and NSF OCI-105357510.1016/j.future.2014.10.008.
26. Sperhac JM, DeLeon RL, White JP, et al. Towards performant workflows, monitoring and measuring. *IEEE ICCCN ’20IEEEIEEE Computer Society*. New York, NY: Springer; 2020.
27. U.S. Geological Survey Landsat 9: USGS, Reston, VA; 2019. https://doi.org/10.3133/fs20193008.
28. Landsat science products. https://www.usgs.gov/land-resources/nli/landsat/landsat-science-product%s. Accessed March 20, 2020.
29. ICESat-2 data products. https://icesat-2.gsfc.nasa.gov/science/data-products. Accessed March 20, 2020.

30. NASA Operation ice bridge. https://www.nasa.gov/mission_pages/icebridge/index.html. Accessed April 8, 2020.
31. European Space Agency (ESA) sentinel. https://sentinel.esa.int/web/sentinel/sentinel-data-access. Accessed March 20, 2020.
32. United States Geological Survey (USGS) Earth Explorer https://earthexplorer.usgs.gov. Accessed March 20, 2020.
33. National Snow and Ice Data Center (NSIDC), U.S. https://nsidc.org/. Accessed March 20, 2020.
34. European Space Agency (ESA) Copernicus. https://spacedata.copernicus.eu/. Accessed March 20, 2020.
35. Schenk T, Csatho B. A new methodology for detecting ice sheet surface elevation changes from laser altimetry data. *IEEE Trans Geosci Remote Sens*. 2012;50:3302-3316.
36. Schenk T, Csatho B, van der Veen C, McCormick D. Fusion of multi-sensor surface elevation data for improved characterization of rapidly changing outlet glaciers in Greenland. *Remote Sens Environ*. 2014;149:239-251.
37. Pollard David, DeConto Robert M. Modelling West Antarctic ice sheet growth and collapse through the past five million years. *Nature*. 2009;458(7236):329–332. 10.1038/nature07809.
38. Dalton AS, Margold M, Stokes CR, et al. An updated radiocarbon-based ice margin chronology for the last deglaciation of the North American Ice Sheet Complex. *Quat Sci Rev*. 2020;234:1–27. https://doi.org/10.1016/j.quascirev.2020.106223.
39. Hughes ALC, Gyllencreutz R, Lohne ØS, Mangerud J, Svendsen JI. The last Eurasian ice sheets – a chronological database and time-slice reconstruction, DATED-1. *Boreas*. 2016;45(1):1-45. https://doi.org/10.1111/bor.12142..
40. Herring TA, Melbourne TI, Murray MH, et al. Plate boundary observatory and related networks: GPS data analysis methods and geodetic products. *Rev Geophys*. 2016;54(4):759-808. https://doi.org/10.1002/2016RG000529.
41. Bevis M, Wahr J, Khan SA, et al. Bedrock displacements in Greenland manifest ice mass variations, climate cycles and climate change. *Proc Natl Acad Sci*. 2012;109(30):11944-11948. https://doi.org/10.1073/pnas.1204664109.
42. Shepherd A, Ivins E, Aa G, et al. A reconciled estimate of ice-sheet mass balance. *Science*. 2012;338:1183-1189. https://doi.org/10.1126/science.1228102.
43. Shepherd A, Ivins E, Rignot E, et al. Mass balance of the Antarctic ice sheet from 1992 to 2017. *Nature*. 2018;558:219-222. https://doi.org/10.1038/s41586-018-0179-y.
44. Shepherd A, Ivins E, Rignot E, et al. Mass balance of the Greenland ice sheet from 1992 to 2018. *Nature*. 2020;579(7798):233-239. https://doi.org/10.1038/s41586-019-1855-2.
45. Nowicki SMJ, Payne A, Larour E, et al. Ice sheet model intercomparison project (ISMIP6) contribution to CMIP6. *Geosci Model Dev*. 2016;9(12):4521-4545. https://doi.org/10.5194/gmd-9-4521-2016.
46. Forster RR, Box JE, van den Broeke MR, Miège C, Burgess EW, van Angelen JH. Extensive liquid meltwater storage in firn within the Greenland ice sheet. *Nat Geosci*. 2014;7:95-98. https://doi.org/10.1038/ngeo2043.
47. Poinar K, Joughin I, Lilien D, Brucker L, Kehrl L, Nowicki S. Drainage of Southeast Greenland Firn aquifer water through crevasses to the bed. *Front Earth Sci*. 2017;5:1–15. https://doi.org/10.3389/feart.2017.00005.
48. Poinar K, Dow CF, Andrews LC. Long-term support of an active subglacial hydrologic system in southeast Greenland by Firn aquifers. *Geophys Res Lett*. 2019;46(9):4772-4781.
49. Miège C, Forster RR, Brucker L, et al. Spatial extent and temporal variability of Greenland firn aquifers detected by ground and airborne radars. *J Geophys Res Earth*. 2016;121(12):2381-2398.
50. Petty AA, Tsamados MC, Kurtz NT, et al. Characterizing Arctic sea ice topography using high-resolution IceBridge data. *Cryosphere*. 2016;10:1161-1179. https://doi.org/10.5194/tc-10-1161-2016.
51. Martin CF, Krabill WIB, Manizade SS, et al. *Airborne Topographic Mapper Calibration Procedures and Accuracy Assessment*. Greenbelt, MD: NASA; 2012.
52. Yoo AB, Jette MA, Grondona M. SLURM: simple linux utility for resource management. In: Feitelson D, Rudolph L, Schwiegelshohn U, eds. *Lecture Notes in Computer Science*. Berlin, Germany: Springer; 2003.
53. Studinger M.. IceBridge ATM L1B Elevation and Return Strength, Version 2. Boulder, Colorado: NASA National Snow and Ice Data Center Distributed Active Archive Center 2013; 2020. https://doi.org/10.5067/19SIM5TXKPGT. Accessed March 20, 2020.
54. Cicero EC, Poinar K, Jones-Ivey R, Sperhac JM, Petty A, Briner JP. Distribution and evolution of crevasses draining the firn aquifer in southeast Greenland identified with ghub. American Geophysical Union no. 737787; 2020.
55. Goelzer H, Nowicki S, Edwards T, et al. Design and results of the ice sheet model initialisation experiments initMIP-Greenland: an ISMIP6 intercomparison. *Cryosphere*. 2018;12(4):1433-1460. https://doi.org/10.5194/tc-12-1433-2018.
56. Seroussi H, Nowicki S, Simon E, et al. initMIP-Antarctica: an ice sheet model initialization experiment of ISMIP6. *Cryosphere*. 2019;13(5):1441-1471. https://doi.org/10.5194/tc-13-1441-2019.
57. Taylor KE. Summarizing multiple aspects of model performance in a single diagram. *J Geophys Res Atmos*. 2001;106(D7):7183-7192. https://doi.org/10.1029/2000JD900719.
58. Eyring V, Righi M, Lauer A, et al. ESMValTool (v1.0) – a community diagnostic and performance metrics tool for routine evaluation of Earth system models in CMIP. *Geosci Model Dev*. 2016;9(5):1747-1802. https://doi.org/10.5194/gmd-9-1747-2016.
59. Barthel A, Agosta C, Little CM, et al. CMIP5 model selection for ISMIP6 ice sheet model forcing: Greenland and Antarctica. *Cryosphere*. 2020;14(3):855-879. https://doi.org/10.5194/tc-14-855-2020.
60. Slater DA, Straneo F, Felikson D, et al. Estimating Greenland tidewater glacier retreat driven by submarine melting. *Cryosphere*. 2019;13(9):2489-2509. https://doi.org/10.5194/tc-13-2489-2019.
61. Slater DA, Felikson D, Straneo F, et al. Twenty-first century ocean forcing of the Greenland ice sheet for modelling of sea level contribution. *Cryosphere*. 2020;14(3):985-1008. https://doi.org/10.5194/tc-14-985-2020.
62. NASA Goddard Laboratory Cryosphere Model Comparison Tool (CmCt) tool https://ggsghpcc.sgt-inc.com/cmct/. Accessed March 12, 2020.
63. Price SF, Hoffman MJ, Bonin JA, et al. An ice sheet model validation framework for the Greenland ice sheet. *Geosci Model Dev*. 2017;10(1):255-270. https://doi.org/10.5194/gmd-10-255-2017.
64. NASA Goddard space flight center Ice, Cloud, and land Elevation Satellite (ICEsat); 2009. https://icesat.gsfc.nasa.gov.
65. NASA Jet Propulsion Laboratory Gravity recovery and climate experiment (GRACE); 2002. https://www.jpl.nasa.gov/missions/gravity-recovery-and-climate-experime%nt-grace.
66. Virtual Earth System Laboratory (VESL). https://vesl.jpl.nasa.gov/. Accessed March 20, 2020.

67. Cheng DLC, Larour EY, Quinn JD, Halkides DJ. VESL: the virtual earth sheet laboratory for ice sheet modeling and visualization. Paper presented at: Proceedings of the AGU Fall Meeting Abstracts; 2016; New York, NY.
68. Larour E, Cheng D, Perez G, et al. A JavaScript API for the ice sheet system model (ISSM) 4.11: towards an online interactive model for the cryosphere community. *Geosci Model Dev*. 2017;10(12):4393-4403. https://doi.org/10.5194/gmd-10-4393-2017.
69. Columbia Glacier model, Virtual Earth System Laboratory (VESL) Library. https://vesl.jpl.nasa.gov/glaciers/columbia/. Accessed March 20, 2020.
70. Shekhar P, Patra A. Hierarchical approximations for data reduction and learning at multiple scales. *Foundations of Data Science*. 2020;2(2):123–154. http://dx.doi.org/10.3934/fods.2020008.
71. Shekhar P, Csatho B, Schenk T, Roberts C, Patra AK. ALPS: A Unified Framework for Modeling Time Series of Land Ice Changes. *IEEE Transactions on Geoscience and Remote Sensing*. 2020;1–16. http://dx.doi.org/10.1109/tgrs.2020.3027190.
72. Center for Computational Research (CCR), University at Buffalo. http://hdl.handle.net/10477/79221. Accessed March 20, 2020.
73. Rew R, Davis G. NetCDF: an interface for scientific data access. *IEEE Comput Graph Appl*. 1990;10(4):76-82.
74. UCAR/Unidata Network Common Data Form (netCDF) Version 4.7.4 Software; 2020. https://doi.org/10.5065/D6H70CW6.

How to cite this article: Sperhac JM, Poinar K, Jones-Ivey R, et al. *GHub*: Building a glaciology gateway to unify a community. *Concurrency Computat Pract Exper*. 2020;e6130. https://doi.org/10.1002/cpe.6130

APPENDIX A. ISMIP6 DATA FORMAT AND DIRECTORY STRUCTURE

The ISMIP6 files are stored on backed-up, secure media at University at Buffalo CCR's machine room. File format is in the common, flexible netCDF format.[73,74] The organization of the ISMIP6 data directories follows a well-understood convention:

```
ghub/ISMIP6/Projections/output/
└─ice sheet location (Greenland or Antarctica)
   └─modeling group name (such as JPL, NCAR)
      └─ice sheet model name (such as IMAUICE)
         └─experiment name and grid size (such as exp10, 05)
```

The experiment name directories, which contain information about the grid size of the files within, contain data files. The ISMIP6 file naming convention encodes the 2D variable captured in a given dataset, the location, modeling group name, ice sheet model name, and experiment name. Example file name:

```
acabf_GIS_IMAU_IMAUICE1_init.nc
```

Where:

- acabf—2D variable name
- GIS—Greenland ice sheet
- IMAU—Modeling group name
- IMAUICE1—Ice sheet model name
- init—experiment name