



Annual User Assessment - 2019 Summary Report

supported by

*National Science Foundation
Office of Advanced Cyberinfrastructure
Award #1445604*

prepared for

**David Y. Hancock
Winona Snapp-Childs**

Research Technologies
Indiana University Pervasive Technology Institute

Jetstream Advisory Board

prepared by

**Julie A. Wernert
Harmony E. Jankowski**

Indiana University Pervasive Technology Institute

November 2019
Version 1.1

Table of Contents

A.	Document History	3
B.	Document Scope and Findings	4
B.1.	Findings	4
C.	Survey Methodology and Results	6
C.1.	Methodology	6
C.2.	User Satisfaction with Jetstream.....	6
C.3.	Year-to-Year Satisfaction with Jetstream Services	8
C.4.	Importance of Jetstream to Research and Educational Activities.....	9
C.5.	Storage Capacity Limitations.....	10
C.6.	Unique Features of Jetstream.....	11
C.7.	Commercial Cloud Usage.....	12
C.8.	Training	13
D.	Appendix	14
D.1.	Appendix Part 1 - Final Questionnaire.....	14
D.2.	Appendix Part 2 — Open text responses for: Please tell us more about your dissatisfaction with Jetstream services and/or resources.....	31
D.3.	Appendix Part 3 — Text Comments for: Please explain how the limitation on the capacity of Jetstream’s storage allocations has impacted your ability to perform research.....	33
D.4.	Appendix Part 4 — Text Comments for: Considering your use of Jetstream relative to other national resources (e.g., UCSD, Comet, PSC, Bridges, TACC, Stampede 2, etc.), which of its unique features are most useful to your research program? Other:	35
D.5.	Appendix Part 5 — Text Comments for: Please share with us any additional information you feel necessary to understand your storage needs.	35
D.6.	Appendix Part 6 — Text Comments for: For which activities are you using commercial cloud resources?.....	36
D.7.	Appendix Part 7 — Text Comments for: Please share with us any general comments about your experiences with Jetstream. (For example, are there any current features that you find particularly useful? How has the use of Jetstream improved your ability to perform research? Etc.)	36
D.8.	Appendix Part 8 — Text Comments for: Please provide citations for the products that you indicated have benefitted from the use of Jetstream, including title, authors, publication, publication date, when and where presented, URL, and/or digital object identifier (DOI).....	41

A. Document History

Relevant Sections	Version	Date	Changes	Author
Entire Document	1.00	10/11/2019	Document creation	H. Jankowski
Appendices	1.00	11/12/2019	Appendices added	H. Jankowski
Entire document	1.00	11/27/2019	Analysis tables and narrative; findings summary	J. Wernert
Entire document	1.00	1/8/2020	Full copyedit; formatting adjustments	H. Jankowski
Findings section	1.1	1/15/2020	Addressing PI comments	J Wernert

B. Document Scope and Findings

This document summarizes the results of the Jetstream Annual User Assessment conducted among 635 users from September 11 through November 5, 2019. This was the fourth survey of Jetstream users. It was designed primarily as a tool to gauge broad, overall satisfaction with Jetstream activities and services — a basic “report card.” Other surveys and feedback mechanisms are deployed throughout the project to gather more granular, project-specific, and/or resource- and service-specific data.

As with any survey instrument or resulting report, one should exercise caution in inferring too much with regard to specific results, either positive or negative. The authors made every effort to summarize and convey the survey results accurately and as received so as to not introduce any bias. Nonetheless, readers should pay specific attention to the data collection methodology, especially sources of survey error, detailed in Section C.

Please direct any questions regarding the methods used in the administration of this survey and/or the summarizing of responses provided in this report to Julie Wernert at the Indiana University Pervasive Technology Institute, jwernert@iu.edu.

B.1. Findings

This report provides a basic analysis of the 2019 Jetstream Annual User Assessment. Section C describes the survey’s data collection methodology. The survey consisted of quantitative and qualitative questions designed to determine user satisfaction of Jetstream services and resources.

Findings from the Jetstream Annual User Assessment include the following:

- Overall, over nearly 88 percent of all respondents report they are either “satisfied” or “extremely satisfied” in their experience using the Jetstream system. Mean satisfaction with Jetstream is 4.30 (on a 5.0 scale). Satisfaction was down slightly from 2018.
- With the exception of documentation, all evaluated service areas scored above 4.0 (on a 5.0 scale.)
- Users note the highest levels of satisfaction with the quality of responses to questions, whether via help@xsede.org (4.53/5.0) or via direct email to Jetstream staff (4.51/5.0), followed by the speed of response to questions via direct email to Jetstream staff and/or help@xsede.org (4.48/5.0) and the quality of in-person training and workshops (4.35).
- Some 74 percent of respondents report that Jetstream is “very important” or “essential” to their research activities with a mean importance rating of 4.14 (on a 5.0 scale).
- Just over 73 percent of users indicate the resource is “very important” or “essential” to their education activities, with a mean importance rating of 4.00 (on a 5.0 scale).
- When asked if Jetstream’s limited storage allocation impacted the ability to conduct research fully, twenty-two percent of those responding to this question (42/190) responded in the affirmative, indicating that their research was impacted by this constraint. Some 57 percent of these users indicated that a 10-25TB allocation would best satisfy their needs, while another 19 percent said 25-50TB would be needed to accommodate their research programs.

- The ability to create custom VMs was cited as Jetstream’s most useful feature by 36 percent of respondents, followed by some 35 percent of users who cited on-demand ability to access computational resources as most useful.
- Of those responding about their use of commercial cloud resources, 43 percent (81/189) indicate they use commercial cloud resources for at least some of their professional activities; of these, half report using Amazon Web Services (AWS) and another quarter using the Google Cloud Platform.
- Some 60 percent (55/92) of commercial cloud users indicate that they use both Jetstream and commercial cloud resources to conduct their research. Nearly eight percent of users (7/92) report having moved from the commercial cloud to Jetstream, and some 14 percent (13/92) have used Jetstream to pilot work before scaling up.
- Users were largely neutral-to-positive about the training methods they were asked to rate, but showed a clear preference for the ability to self-serve with just-in-time, online resources. In particular, web documentation and self-paced online tutorials are preferred by Jetstream users.
- Users from minority-serving institutions and/or EPSCoR states comprise only five percent of the Jetstream user population.
- The most typical Jetstream user is male, caucasian, and of non-Hispanic or Latino ethnicity, working at a doctoral-granting and/or research institution in the biology and/or computer and information science fields.

C. Survey Methodology and Results

C.1. Methodology

The Jetstream Annual User Assessment was conducted from September 11 through November 5, 2019, and was available to all Jetstream users who had accessed Jetstream resources in the preceding twelve months, excepting student users and those with training accounts; an additional panel of users who had participated in the 2018 user survey was included for long-term longitudinal tracking. (The survey questionnaire is included in Section D.1.) Users were not required to participate and were able to opt out of the survey. Those who chose to participate were only eligible to do so once during the deployment period; the survey was accessed via a custom link specific to each member of the population and after completing the survey (or opting out), the survey was no longer accessible. Users were sent an initial letter of invitation via email, followed by up to five reminder messages sent to non-respondents and partials (those who had started the survey but had not yet completed it).

The total number of users in the population was 635 (excluding those users whose invitation was returned as undeliverable due to outdated contact information), with 216 choosing to participate at some level. Respondents were not required to answer any questions and could skip questions they did not wish to answer and, therefore, the N varies from question to question. The effective rate of response was 34%.








The data collection instrument was submitted (under protocol #1703859488A002) to the Indiana University Institutional Review Board, and was granted “exempt” status. As such, the resulting data may be published externally in reports, presentations, and other documents, so long as data are presented in aggregate form and no identifying information is divulged. While all identifying information is redacted for public consumption, please note that responses were initially associated with the respondent’s email address and cannot be considered anonymous; in some cases, this may have a bearing on the responses provided and should be considered when reviewing and analyzing the data.

Surveys of this kind are sometimes subject to types of inaccuracies for which precise estimates cannot be calculated. For example, findings may be influenced by events that take place while the survey is in the field. Events occurring since the time the surveys were completed could have changed the opinions reported here. Sometimes questions are inadvertently biased or misleading. The views of people who responded to the survey may not necessarily replicate the views of those who did not respond to the survey.

C.2. User Satisfaction with Jetstream


In this survey, users were asked to rate their levels of satisfaction, ranging from “extremely dissatisfied” to “extremely satisfied,” with Jetstream services in 10 areas (Table 1), as well as their overall satisfaction with Jetstream (Table 2). With the exception of Jetstream documentation, all evaluated areas scored above 4.0 (on a 5.0 scale.) Users note the highest levels of satisfaction with the quality of response to questions, whether via help@xsede.org (4.53/5.0) or via direct email to Jetstream staff (4.51/5.0), followed by the speed of response to questions via direct email to Jetstream staff and/or help@xsede.org (4.48/5.0) and the quality of in-person training and workshops (4.35).

Table 1. Satisfaction with Jetstream services

Please rate your satisfaction with the following aspects of Jetstream on a scale of 1 to 5, with 1 being "extremely dissatisfied" and 5 being "extremely satisfied." If you have no basis for rating your satisfaction, please select "Not applicable."									
	Mean Satisfaction	Number of Applicable Responses	Distribution (1 = extremely dissatisfied, 5 = extremely satisfied)					Number providing no response	Histogram
			1	2	3	4	5		
Availability of VM images to solve my problems	4.27	181	0.6%	3.3%	8.8%	43.6%	43.65%	35	
Speed (responsiveness) of Jetstream	4.15	190	0.5%	5.8%	12.6%	40.5%	40.53%	26	
Documentation about Jetstream	3.97	186	0.5%	5.4%	21.5%	41.9%	30.65%	30	
Atmosphere, the Jetstream User Portal	4.09	175	2.3%	4.6%	10.3%	47.4%	35.43%	41	
Speed of response to my questions via help@xsede.org	4.48	160	1.3%	2.5%	3.8%	32.5%	60.00%	56	
Quality of response to my questions via help@xsede.org	4.53	158	1.3%	1.3%	4.4%	29.1%	63.92%	58	
Speed of response to my questions via direct email to Jetstream staff?	4.48	153	1.3%	0.7%	7.8%	28.8%	61.44%	63	
Quality of response to my questions via direct email to Jetstream staff?	4.51	155	1.3%	0.6%	7.1%	27.7%	63.23%	61	
Quality of in-person workshops and training	4.35	77	0.0%	3.9%	13.0%	27.3%	55.84%	139	
Quality of online workshops and training	4.22	90	0.0%	5.6%	16.7%	27.8%	50.00%	126	

Overall, nearly 88 percent of all respondents report they are either "satisfied" or "extremely satisfied" with their experience using the Jetstream system. This is down slightly from 2018, when some 90 percent reported they were "satisfied" or "extremely satisfied." Applying a standard Likert scale to the responses offered, with "1" being "extremely dissatisfied" and "5" being "extremely satisfied," the mean satisfaction is 4.30 on a 5.0 scale. Overall mean satisfaction also dipped from its peak of 4.32, first achieved in 2017. The mean satisfaction and rating distribution are presented in Table 2.

Table 2. Overall satisfaction with Jetstream

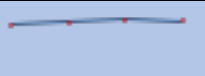






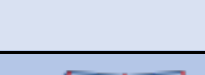



Please rate your satisfaction with the following aspects of Jetstream on a scale of 1 to 5, with 1 being "extremely dissatisfied" and 5 being "extremely satisfied." If you have no basis for rating your satisfaction, please select "Not applicable."									
	Mean Satisfaction	Number of Applicable Responses	Distribution (1 = extremely dissatisfied, 5 = extremely satisfied)					Number providing no response	Histogram
			1	2	3	4	5		
Overall performance of Jetstream	4.30	188	0.0%	3.7%	8.5%	41.5%	46.3%	28	

Respondents who indicated they were either “extremely dissatisfied” or “dissatisfied” were asked to indicate the reasons for their dissatisfaction. Respondents cited issues related to difficulties shelving and unshelving instances; limited and/or confusing documentation; inability to identify unique instances in a different context; the time it takes to set up/deploy instances; sub-par status and progress feedback; the tendency for the Atmosphere interface to be “buggy” and “slow to update,” and the propensity for OpenStack APIs to fall over when attempting to scale up through common automation tooling, among others, as reasons for dissatisfaction. All open text responses to this question are included in Appendix D.2.

C.3. Year-to-Year Satisfaction with Jetstream Services

Since 2016, user satisfaction with various Jetstream services has been assessed. While there have been modest increases and decreases over the four-year period, satisfaction in all areas has remained high, with most areas maintaining mean satisfaction scores at or above 4.0 on a 5.0 scale (Table 3). Notably, the resources garnering the highest and lowest levels of satisfaction have remained largely consistent from year to year, with the speed and quality of help responses (whether via xsede.org or direct interaction with Jetstream staff) receiving the highest levels of satisfaction, and documentation resources receiving among the lowest satisfaction ratings.



Table 3. Mean satisfaction Year-to-Year Trend (2016-2019)

Mean Satisfaction Year-to-Year Trend (2016-2019)					
Service	2016	2017	2018	2019	Annual Trend (range 3.84-4.70)
Availability of VM images to solve my problems	4.02	4.17	4.34	4.27	
Speed (responsiveness) of Jetstream	4.02	4.19	4.19	4.15	
Documentation about Jetstream	3.84	3.96	4.06	3.97	
Atmosphere, the Jetstream User Portal	4.09	3.94	4.05	4.09	
Speed of response to my questions via help@xsede.org	4.55	4.58	4.47	4.48	
Quality of response to my questions via help@xsede.org	4.60	4.48	4.47	4.53	
Speed of response to my questions via direct email to Jetstream staff?	4.64	4.70	4.51	4.48	
Quality of response to my questions via direct email to Jetstream staff?	4.64	4.70	4.55	4.51	
Quality of in-person workshops and training	*	4.41	4.12	4.35	
Quality of online workshops and training	*	4.16	4.03	4.22	
Overall performance of Jetstream	4.21	4.32	4.32	4.30	

C.4. Importance of Jetstream to Research and Educational Activities

Respondents were also asked to indicate how important the Jetstream system is to their respective research and education activities (Table 4). Some 74 percent of respondents, up from 67 percent in 2018, indicate that Jetstream is “very important” or “essential” to their research activities, with a mean importance rating of 4.14 (on a 5.0 scale). Some 74 percent of respondents, up from 60 percent in 2018, indicate the resource is “very important” or “essential” to their education activities, with a mean importance rating of 4.0 (on a 5.0 scale).

Table 4. Importance of Jetstream to users' research and educational activities

Importance of Jetstream to research and educational activities on a scale of 1-5, with 1 being "not important at all" and 5 being "essential."									
	Mean importance	Number of Applicable Responses	Distribution (1 = not important at all, 5 = essential)					Number providing no response	Histogram
			1	2	3	4	5		
Research Activities	4.14	175	1.7%	8.6%	15.4%	22.3%	52.0%	41	
Education Activities	4.00	152	3.3%	14.5%	9.2%	25.0%	48.03%	64	

Since the first Jetstream User Assessment was conducted in 2016, Jetstream's importance to users' research activities has increased steadily (Table 5). In 2016, just under 28 percent of users indicated Jetstream was essential to their research activities; in 2019, 52 percent indicate the Jetstream system is essential to their research activities, with the most significant increase occurring between 2018 and 2019.

The trajectory of the importance of Jetstream to users' educational activities is less straightforward. In 2016, over 35 percent indicated Jetstream was essential to their educational activities. In 2017 and 2018, this number waned, with just 19.6 and 27.2 percent, respectively, reporting Jetstream as essential. In 2019, 48 percent of users report Jetstream is essential to their educational activities, rebounding to levels that are more in line with what might be expected.

Table 5. Importance of Jetstream to users' research and educational activities (2016-2019)

Importance of Jetstream in conducting your research and educational activities (1=Not at all important, 5=Essential)										
Year		1	2	3	4	5	N	Mean	% of N=4 or 5	% of N=5
2016	Research	10.6%	2.1%	19.1%	40.4%	27.7%	47	3.72	68.1%	27.7%
	Education	9.5%	7.1%	21.4%	26.2%	35.7%	42	3.71	61.9%	35.7%
2017	Research	6.3%	3.1%	18.9%	40.2%	31.5%	127	3.87	71.7%	31.5%
	Education	7.5%	9.3%	30.8%	32.7%	19.6%	107	3.48	52.3%	19.6%
2018	Research	3.7%	5.1%	24.5%	33.3%	33.3%	216	3.88	66.7%	33.3%
	Education	7.7%	10.1%	22.5%	32.5%	27.2%	169	3.62	59.8%	27.2%
2019	Research	1.7%	8.6%	15.4%	22.3%	52.0%	175	4.14	74.3%	52.0%
	Education	3.3%	14.5%	9.2%	25.0%	48.0%	152	4.00	73.0%	48.0%

C.5. Storage Capacity Limitations

Jetstream presently limits the capacity of its default storage allocation to 1TB, which may impact the ability of some users to conduct their research. With this as context, users were asked if this constraint did, in fact, impact their ability to carry out their research programs. Twenty-two percent of those responding to this question (42/190), responded in the affirmative, indicating that their research was impacted by this constraint. When asked to explain how the limitation on the capacity of Jetstream's storage allocations has impacted the ability to perform research, respondents offered the following comments, among others:

- I work with databases, which require different configurations than what is usually available. Adding extra volumes does not allow a database to expand its storage.
- The default storage associated with a VM is somewhat limited. Although I can connect external volume to a VM, Globus doesn't seem to work with external volume.
- Limited storage capacity has made it difficult to index large corpora and datasets

- The indexing files we generate will approximately take up to a few TB of space, but each VM is only having a few hundred GBs. Perhaps we are not using the correct path for large file storage?

All open text comments in response to this question are included in Appendix D3.

Further, those who responded affirmatively to the question about capacity were asked what capacity range would better satisfy their needs (Table 6.) Some 57 percent indicated that a 10-25TB allocation would satisfy their needs, while another 19 percent said 25-50TB would best accommodate their research programs.

Table 6. Capacity ranges to best satisfy Jetstream users' needs

What capacity range would better satisfy your needs? (N=42)		
10-25TB	24	57%
26-50TB	8	19%
51-100TB	4	10%
>100TB	4	10%
No answer	2	5%

C.6. Unique Features of Jetstream

Respondents were also asked to consider their use of Jetstream within the context of other national resources and identify those unique features that were of most use to their research programs (Table 7). The ability to create custom VMs was cited as most useful by 36 percent of respondents, followed by some 35 percent of users who cited on-demand access to computational resources as most useful. Another 28 percent reported that the Atmosphere user interface was Jetstream's most useful feature in conducting their research. Only 11 percent of respondents cited Jetstream's elastic computing features as most useful.

Table 7. Unique features of Jetstream most useful to users' research programs

Considering your use of Jetstream relative to other national resources (e.g., UCSD, Comet, PSC, Bridges, TACC, Stampede 2, etc.), which of its unique features are most useful to your research program? Select all that apply. (N=309)		
Ability to create custom VMs	111	36%
On-demand ability to access computational resources	109	35%
Atmosphere user interface (i.e., web shell, web desktop)	85	28%
Ability to interactively use software applications, such as RStudio, Matlab, etc.	48	16%
Availability of featured images (e.g., R, Jupyter, Matlab, etc.) beyond standard operating systems	47	15%
Persistent services (e.g., science gateways)	43	14%
Ability to automate VM creation via an API	41	13%
Elastic computing (e.g., Kubernetes, virtual clustering, etc.)	34	11%
Other	11	4%

Other responses include:

- Ease of installing software packages (much simpler on instances with root access than on shared HPC clusters)
- Ability to create and run customized Docker containers for Jupyter Notebook
- Ability to user Rstudio and install new R packages (eg sparklyr for Spark w/ R)
- Ability to be root on the custom VMs
- Make it like AWS
- Ability to allow students interact with Jupyter notebooks directly from their browser without need to install any software
- Haven't used any national resources

- I haven't used national resources
- These are all potentially great but I struggled with getting VMs to load without error

Full-text responses for this question are available in Appendix D.4.

C.7. Commercial Cloud Usage

Jetstream users were also asked about the use of commercial cloud resources in their research and educational programs (Table 8). When asked if they, or members of their team, currently use any commercial cloud resources in their research, scholarship, creative activity, or outreach, nearly 43% of those who responded (81/189) did so in the affirmative, indicating that they presently use commercial cloud resources for at least some portion of their professional activities. Further, those using commercial cloud resources were asked which providers they were currently using, with half of respondents (66/133) reporting they used Amazon Web Services (AWS) and another quarter (33/133) the Google Cloud Platform.

Table 8. Commercial cloud resources used by Jetstream users

Which commercial cloud resources are you currently using? Select all that apply (N=133)		
Amazon Web Services (AWS)	66	50%
Google Cloud Platform (Do not select if only using Google Docs and/or Google Drive)	33	25%
Microsoft Azure	22	17%
IBM Cloud Services	2	2%
Other	10	8%

Other responses included:

- Other: Digital Ocean (4)
- Other: Linode
- Other: Heroku
- Other: For our data processing (InSAR) some colleagues use AWS and have very good experience with it. I was hoping to get the same experience with Jetstream but I (almost) came to the conclusion that Jetstream is not made for our data processing requirements. So I just plan to use it for a Science Gateway.
- Other: Backblaze B2 (only for data backup)
- Other: custom
- Other: My current project only uses Jetstream, on another project my colleagues use Digital Ocean and AWS

Finally, those who use commercial cloud resources were asked to describe how they were using these resources in conjunction with Jetstream (Table 9). Some 60 percent (55/92) indicate that they use both Jetstream and commercial cloud resources to conduct their research. Nearly eight percent of users (7/92)

report having moved from the commercial cloud to Jetstream, and some 14 percent (13/92) have used Jetstream to pilot work before scaling up.






Table 9. Commercial cloud resource usage in conjunction with Jetstream

For which activities are you using commercial cloud resources? <i>Select all that apply.</i> (N=92)		
I used Jetstream to pilot my work before scaling up	13	14.1%
I moved from the commercial cloud to Jetstream	7	7.6%
I use both the commercial cloud and Jetstream for my research	55	59.8%
Other	17	18.5%

C.8. Training

Respondents were asked to consider their use of Jetstream, and then to rate their preferred training delivery formats. Similar to previous years' results, users were largely neutral-to-positive about the training methods they were asked to rate, but showed a clear preference for the ability to self-serve through the use of just-in-time, online resources. Table 10 presents Jetstream users' preferred training methods.

Table 10. Respondents' preferred training methods

Considering your use of Jetstream, please rate your preferences for training delivery on scale of 1-5, with 1 being "strongly do not prefer" and 5 being "strongly prefer."									
	Mean Preference	Number of Applicable Responses	Distribution (1 = strongly do not prefer, 5 = strongly prefer)					Number providing no response	Histogram
			1	2	3	4	5		
Web documentation	4.34	184	1.6%	0.5%	10.9%	36.4%	50.54%	32	
Live, in-person tutorials or workshops	3.32	183	6.0%	13.1%	38.3%	27.9%	14.75%	33	
Live, online webinars (tutorials, workshops, etc.)	3.44	185	7.0%	11.9%	28.6%	34.6%	17.84%	31	
Recording of webinars (with minimal editing)	3.64	188	4.8%	6.9%	30.9%	34.6%	22.87%	28	
High-quality training videos	3.88	187	2.7%	3.7%	23.0%	43.9%	26.74%	29	
Self-paced, online tutorials	4.20	189	0.5%	1.6%	15.3%	42.9%	39.68%	27	

When examining training preferences by population type and/or role, preferences map closely to those of the overall population, with Web documentation and self-paced online tutorials being the most preferred methods. Data does not suggest that any particular field of study or professional role overwhelmingly affects one's preferred method of training delivery.

D. Appendix

D.1. Appendix Part 1 - Final Questionnaire

Jetstream Annual User Survey - 2019

Start of Block: Block 1 - Informed Consent

Jetstream Annual User Assessment – Informed Consent Indiana University Institutional Review Board Protocol #1408987944A004/Exempt

You are invited to participate in the Jetstream Annual User Assessment conducted by principal investigators of the National Science Foundation-funded Jetstream project. We ask that you read this statement and ask any questions you may have before agreeing to take part in the assessment. This assessment is administered on behalf of the Jetstream project by the Indiana University Pervasive Technology Institute and is funded, in part, by the National Science Foundation.

PURPOSE: The purpose of this assessment is aimed at determining current levels of usage and satisfaction with the Jetstream environment and its associated resources and services (e.g., training, support, etc.).

Survey information will be used to improve and expand the services provided by Jetstream and to aid in the decision-making processes related to resource allocation and service expansion and improvements. Survey results may also be used to inform scholarly publications, presentations, and/or funding proposals.

PROCEDURES FOR THE STUDY: If you agree to participate, you will complete an online survey, for which there is no compensation. You will receive via email an initial letter of invitation, followed by up to four (4) reminder messages. After the initial letter of invitation, only those who have not responded will receive subsequent messages. You will have the opportunity to opt out of all future communications upon receipt of the initial letter of invitation. You will have the option to submit for follow up contact. Future contact may be in the form(s) of telephone call, video-conference, in-person interview, and/or focus group. The survey should not take more than 10 minutes to complete.

CONFIDENTIALITY: Every effort will be made to keep any personal information that you inadvertently disclose, as well as project data used to identify population members, confidential. All survey results will be reported in the aggregate and your identity will be held in confidence in reports in which the survey results may be published and/or in databases in which results are stored. Should the resulting data set be made public, it will be redacted of all identifying information. Archived data will be redacted of all identifying information and stored on secure Indiana University systems. However, we cannot guarantee absolute confidentiality. Your personal information may be disclosed if required by law. Organizations that may inspect and/or copy survey records for quality assurance and data analysis include groups such as the study investigator and his/her research associates, the Indiana University Institutional Review Board or its designees, and (as allowed by law) state or federal agencies, specifically the Office for Human Research Protections (OHRP).

CONTACTS FOR QUESTIONS OR PROBLEMS: For questions about this assessment, contact Indiana University Information Manager Julie Wernert at (812) 856-5517 or jwernert@iu.edu. For questions about your rights as a participant or to discuss problems, complaints, or concerns about the assessment; to obtain information, or to offer input, please contact the IU Human Subjects Office at (812) 856-4242 or by email at irb@iu.edu.

VOLUNTARY NATURE OF STUDY: Taking part in this assessment is voluntary. You may choose not to participate, to skip any questions you do not wish to answer, and/or to leave the assessment at any time. Leaving the assessment will not result in any penalty. Your decision whether or not to participate in this assessment will not affect your current or future relations with the Jetstream Project, the Pervasive Technology Institute, Indiana University, or the National Science Foundation. This study was approved by the Indiana University Institutional Review Board on September 9, 2019. Please reference protocol #1408987944A004/Exempt when inquiring.

Do you agree to participate?

Yes (1)

No (2)

*Skip To: End of Block If Jetstream Annual User Assessment – Informed Consent Indiana University Institutional Review Boar...
= Yes*

*Skip To: End of Survey If Jetstream Annual User Assessment – Informed Consent Indiana University Institutional Review Boar...
= No*

End of Block: Block 1 - Informed Consent

Start of Block: Usage Screener

In the past 12 months, **approximately** how many times have you used the Jetstream system?

- I have not used Jetstream in the past 12 months
- 1-5
- 6-10
- 11-20
- 21-50
- 51-100
- More than 100

End of Block: Usage Screener

Start of Block: Non Usage Question

Display This Question:

If In the past 12 months, approximately how many times have you used the Jetstream system? = I have not used Jetstream in the past 12 months

Is there someone else in your group who would be more appropriate for this annual assessment?

- Name _____
- Email _____

Skip To: End of Survey If Is there someone else in your group who would be more appropriate for this annual assessment? Is Displayed

Skip To: End of Survey If Is there someone else in your group who would be more appropriate for this annual assessment? = Name

End of Block: Non Usage Question

Start of Block: Jetstream Usage

In which areas do you use the Jetstream system? *Select all that apply.*

Research

Teaching

Training

Other: _____

Which of the listed Jetstream services do you **primarily** use?

Atmosphere, the Jetstream User Portal

Jetstream API services based on OpenStack

Jetstream OpenStack Horizon Portal

Not sure

Page Break

Please rate your satisfaction with the following aspects of Jetstream on a scale of 1 to 5, with 1 being “extremely dissatisfied” and 5 being “extremely satisfied.” *If you have no basis for rating your satisfaction, please select “Not applicable.”*

	1- Extremely dissatisfied	2- Dissatisfied	3-Neither satisfied or dissatisfied	4-Satisfied	5- Extremely satisfied	X-Not applicable
Availability of VM images to solve my problems	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Speed (responsiveness) of Jetstream	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Documentation about Jetstream	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Atmosphere, the Jetstream User Portal	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Speed of response to my questions via help@xsede.org	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Quality of response to my questions via help@xsede.org	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Speed of response to my questions via direct email to Jetstream staff?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Quality of response to my questions via direct email to Jetstream staff?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Quality of in-person workshops and training	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Quality of online workshops and training	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Overall performance of Jetstream	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Display This Question:

If Please rate your satisfaction with the following aspects of Jetstream on a scale of 1 to 5, wit... = 1-Extremely dissatisfied

Or Please rate your satisfaction with the following aspects of Jetstream on a scale of 1 to 5, wit... = 2-Dissatisfied

Please tell us more about your dissatisfaction with Jetstream services and/or resources.

Page Break

Considering your use of Jetstream relative to other national resources (e.g., UCSD's Comet, PSC's Bridges, TACC's Stampede 2, etc.), which of its unique features are **most useful** to your research program? *Select all that apply.*

- Atmosphere user interface (i.e., web shell, web desktop)
- On-demand ability to access computational resources
- Ability to interactively use software applications, such as RStudio, Matlab, etc.
- Ability to create custom VMs
- Ability to automate VM creation via an API
- Persistent services (e.g., science gateways)
- Elastic computing (e.g., Kubernetes, virtual clustering, etc.)
- Availability of featured images (e.g., R, Jupyter, Matlab, etc.) beyond standard operating systems
- Other: _____

Page Break

End of Block: Jetstream Usage

Start of Block: Importance of Jetstream

Please rate the importance of Jetstream to your **research** activities on a scale of 1-5, with 1 being “not important at all” and 5 being “essential.” If you have no basis for rating Jetstream's importance to your research activities, select “Not applicable.”

- 1-Not important at all
 - 2-Slightly important
 - 3-Moderately important
 - 4-Important
 - 5-Very important
 - Not applicable
-

Please rate the importance of Jetstream to your **educational** activities on a scale of 1-5, with 1 being “not important at all” and 5 being “essential.” If you have no basis for rating Jetstream's importance to your educational activities, select “Not applicable.”

- 1-Not important at all
- 2-Slightly important
- 3-Moderately important
- 4-Important
- 5-Very important
- Not applicable

End of Block: Importance of Jetstream

Start of Block: Training

Considering your use of Jetstream, please rate your preferences for training delivery formats on a scale of 1 to 5, with 1 being “strongly do not prefer” and 5 being “strongly prefer.”

	1-Strongly do not prefer	2-Do not prefer	3-Neutral	4-Prefer	5-Strongly prefer
Web documentation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Live, in-person tutorials/workshops	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Live, online webinars (tutorials, workshops, etc.)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Recording of webinars (with minimal editing)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
High-quality training videos	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Self-paced, online tutorials	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

End of Block: Training

Start of Block: Storage Allocation Capacity Screener

Has the limitation on the capacity of Jetstream’s storage allocations impacted your ability to do your research?

- Yes
- No

End of Block: Storage Allocation Capacity Screener

Start of Block: Storage Allocation Capacity

Display This Question:
If Has the limitation on the capacity of Jetstream’s storage allocations impacted your ability to do... = Yes

Please explain how the limitation on the capacity of Jetstream’s storage allocations has impacted your ability to perform research?

Display This Question:

If Has the limitation on the capacity of Jetstream's storage allocations impacted your ability to do... = Yes

What capacity range would better satisfy your needs?

- 10-25TB
- 26-50TB
- 51-100TB
- >100TB

Display This Question:

If Has the limitation on the capacity of Jetstream's storage allocations impacted your ability to do... = Yes

Please share with us any additional information you feel necessary to understand your storage needs.

End of Block: Storage Allocation Capacity

Start of Block: Commercial Cloud Usage

Do you or members of your team currently use any commercial cloud resources in your research, scholarship, creative activity, or outreach?

- Yes
- No

End of Block: Commercial Cloud Usage

Start of Block: Commercial Cloud Usage

Display This Question:

If Do you or members of your team currently use any commercial cloud resources in your research, sch... = Yes

Which commercial cloud resources are you currently using? *Select all that apply.*

- Amazon Web Services (AWS)
- Google Cloud Platform (Please do not select if you are only using Google Docs and/or Google Drive)
- IBM Cloud Services
- Microsoft Azure
- Other: _____

Display This Question:

If Do you or members of your team currently use any commercial cloud resources in your research, sch... = Yes

For which activities are you using commercial cloud resources? *Select all that apply.*

- I used Jetstream to pilot my work before scaling up
- I moved from the commercial cloud to Jetstream
- I use both the commercial cloud and Jetstream for my research
- Other: _____

End of Block: Commercial Cloud Usage

Start of Block: Open Text

Please share with us any general comments about your experiences with Jetstream. (For example, Are there are current features that you find particularly useful? How has the use of Jetstream improved your ability to perform research? Etc.)

End of Block: Open Text

Start of Block: Block 5 -Citations

Please tell us about any products you have produced that have benefited in some way from use of Jetstream. *Select all that apply.*

- Not applicable
 - Journal articles
 - Books
 - Book Chapters
 - Thesis/Dissertations
 - Conference Papers and Presentations
 - Other Publications
 - Technologies or Techniques
 - Patents
 - Inventions
 - Licenses
 - Datasets
 - Websites
 - Other: _____
-

Display This Question:

If Please tell us about any products you have produced that have benefited in some way from use of J... != Not applicable

Please provide citations for the products that you indicated have benefited from the use of Jetstream, including title, authors, publication, publication date, when and where presented, URL, and/or digital object identifier (DOI).

End of Block: Block 5 -Citations

Start of Block: Block 4 - Contact

If necessary, may we contact you for additional feedback about your Jetstream user experience?

- Yes
- No

Skip To: End of Block If necessary, may we contact you for additional feedback about your Jetstream user experience? = No

If additional follow-up is necessary, may we share your survey responses and comments with the Jetstream principal(s) who will contact you? (Survey responses will not be shared with those who may be contacting you for additional feedback unless you grant permission.)

- Yes
- No

End of Block: Block 4 - Contact

Start of Block: Block 3 - Demographics

Please indicate the **primary** discipline in which you conduct **research** activities.

▼ Arts and Humanities (1) ... Not applicable (19)

Please indicate the **primary** discipline in which you conduct **educational** activities.

▼ Arts and Humanities (1) ... Not applicable (19)

Please select your **primary** role - the one that best describes your work relative to Jetstream cloud environment.

- University faculty or equivalent
- University/Center research staff or equivalent (non-postdoctoral)
- University/Center non-research support staff (or equivalent)
- Postdoctoral fellow
- Graduate student
- Undergraduate student
- XSEDE-funded staff
- XSEDE Campus Champion
- Executive leadership (e.g., dean, director, vice president, CIO, etc.)
- Other: _____

Please describe your institution. *Select all that apply.*

- EPSCoR institution
 - Minority-serving institution
 - Associate's college (all degrees are at the associate's level)
 - Baccalaureate college/university
 - Master's college/university
 - Doctorate-granting university
 - Teaching-focused institution
 - Research-focused institution
 - Government lab or center
 - Non-profit organization (non-academic)
 - Corporate or industrial organization
 - Non-US institution
 - Other: _____
-

What is your gender?

- Male
 - Female
 - Non-cisgender
 - Other: _____
 - Prefer not to disclose
-

What is your ethnicity?

- Hispanic or Latino
 - Not Hispanic or Latino
 - Prefer not to disclose
-

What is your race? *Select all that apply.*

- Asian
- Black or African-American
- Caucasian
- Native American (including Alaska Native)
- Native Hawaiian or Other Pacific Islander
- Other: _____
- Prefer not to disclose

D.2. Appendix Part 2 — Open text responses for: Please tell us more about your dissatisfaction with Jetstream services and/or resources.

- Does not support teaching well - images have to be customized, cannot add users in bulk, no webinars or online training at IU
- I cannot replicate our interactive CFD simulations on the Bridges GPU system. I guess I gave up because I have to teach and it was easier to go back to workstations. The other problem is that our CFD teaching software is on a license server here at school and we require a VPN tunnel from Bridges GPU to our license server. This is the software company's requirement. I'm able to do this with AWS using AD Connector and a private VPN tunnel. I could sure use the horsepower and low cost. Darn...
- Constantly gets stuck in unshelving or reshelving or when creating an instance. While the staff is quick to respond and usually fixes the issue, I would prefer it worked on its own
- I've found the help documentation to be confusing.
- Spooling up and terminating VMs is a rather slow process and does not show any indications of gradual progress, simply a change in states from "shutdown" to "starting."
- I do not know of any Jetstream online workshop. If there are I think that they are not advertised sufficiently.
- Oh boy, I'd say everything. Sometimes is painfully slow. Sometimes instances do not do what the button clicked told them to do. The instances are uniquely recognized just by their UUID, not their name, which I often have to change, but the emails notifying me of things (e.g. shelved instances to be deleted) refer to the UUID and the *first* name used for the instance, which usually is not the one relevant anymore! Moreover, instances cannot be organized in a hierarchy in the web interface, e.g. with folders or sorting them according to some criteria. These latter two problems have forced me to use way less instances that I desired, to avoid the "needle in haystack" issues. Moreover instances go into "error" mode without further explanation....But all this is just to name a few things: there is more....My time for this survey is limited so let me paste below some notes I wrote for a different context (so they may not tailored for this survey and you will notice it) but hopefully it will give you a better idea of my dissatisfaction.
 - [Paste begins] One unfortunate issue is that I cannot uniquely identify instances in different context. If I could probably most of these would be moot point, or I could design workarounds. For example, if the IP numbers were always fixed for my instances, I could make a look-up-table and use the IP# for this purpose, but unfortunately it changes when the instance is "massaged" for example in atmosphere. So the following are the operations I need to do. 1) creation, activation, shutdown, shelving etc.: happens on atmosphere web interface. Main operation for instance management is association between instance IDs and IP numbers, and causes changes in the latter :-(, 2) deployment: happens on a third machine, formerly my laptop, just recently switched to another Jetstream instance. It uses ansible and github. Many things happen here, but the main operation for instance management is associate IP address with DNS name (and therefore function, or what else ansible need to deploy there). Sometimes the deployment takes forever and I would like to shutdown my laptop but I can't.... very annoying (one of several reasons which motivated myself to move to

a VM in Jetstream itself), 3) Domain validation for HTTPS and actual DNS setting: happens on the server itself, manual run. Main operation for instance management is connecting with letsencrypt and google DDNS servers, 4) Final server turn-on and validation (not really an instance management, but...) done manually on the server itself and on a random machine to verify that everything is working correctly (smoke test, if you wish). Problems: a) instance shutdown/shelving unaware of point 3) leaving DNS pointing to IP addresses eventually assigned to other domains, b) letsencrypt unaware of 1) and sending me "certificate is going to expire" emails for instances that are shutdown/shelved (the certificate is anyway useless, because at unshelving/turning on the IP will change and bullet 3 will need to happen again) or even deleted (so there is no certificate whatsoever...) -- because of the format of such emails, it's very hard to track what-is-what, c) since deployment takes long time, sometimes after 2) is started I get on doing other things and forget to track and complete steps 3) and/or 4) leaving the instance in a limbo which is not clear anywhere but my memory (if I remember!!) -- how to track this? I was thinking to add a last step into ansible to send me an email, but not sure if there is something better, d) I have some test deployment on several dedicated test URLs. I need an authoritative way of saying which instance was for what test (possibly from which github branch it had been deployed, for what purpose). Even when the URL is they utilized by something else, and the IP changes because the instance is turned off then on again. My dream: "something" (let's say the atmosphere interface but does not have to be, nor it needs to be on the web) which tells me clearly: - what instances I have (up, down, shelved, etc., perhaps even the deleted ones), - each one of them in what status is (and I don't mean "up or down" that's the previous bullet, but the "setup" status: steps 1-4 above, including expiration date for the CA certificate currently in use, if applicable)

- The OpenStack APIs tend to fall over when attempting to scale up through common automation tooling. I appreciate your help with sharing the valuable resources with us, however, I do find some inconvenience when using the system I wish to talk about here. 1. The images: it seems the images are kind of old and I couldn't find anything like the latest Ubuntu releases. It would be great if we could import docker containers for use, we could create our own images or we could customize and save a template image, 2. The speed: I'm talking about the 1st time set-up here. It takes a really long time before an image is ready to be used, and what makes it worse is that the portal sometimes is showing wrong/outdated status for the VMs. I wish it could be set-up faster and have better VM status updates, 3. The document: it's not too bad, but I feel struggling when my first time using the portal, and especially when the document on xsede website is wrong (it still says we can use the login hub to login to the system, which is not true). While, for a used-to user, these are not problems. It's still a very good system to use.
- The openstack web interface is usually extremely slow.
- The atmosphere ui is bare bones and does not offer the tools I need to run my instances. I had to switch to the Openstack API & Horizon web interface but still have to go back to atmosphere to review resource utilization information.
- No hands on. For example would be nice to how to use Bridge, a tutorial at least, I don't know how to use bridge and the documentation is not friendly, I am more like a tutorial video guy
- I was running genome annotation project using Jetstream. I was promised by one of its researchers that he would be helping me to carry out the analysis. Whenever I ran into some problem, I did not

get enough assistance from him. Hence I would kindly request you to provide assistance to the people who have been provided with allocation to do their project.

- starting the previous project always need a long time and sometimes error occurs as well. Maybe that's because some of my projects are too big but I really hope this part can be fast.
- I find the Atmosphere interface to be buggy, slow to update, and activating VMs is slow and error-prone, without clear indication of how to avoid errors beyond 'try a hard reboot.' In my limited experience, it seems the more complicated the VM or the more resources requested, the greater chance of failure.
- Atmosphere is buggy in that instances don't always come up/seem to get stuck a lot; the interface desperately needs UI design help; the guacamole server's lack of seamless copy-paste is extremely frustrating and the old web shell is no longer an option.
- Horizon dashboard is laggy.
- The web interface is clunky and quite hard to manage (seems based on Amazon; not the best interface there, either). We also had some difficulty accessing VMs in our environment (standard desktop iMacs behind BIDMC's firewall) that took quite a lot of doing to get things working.
- A lot of instability issues preventing our research
- Application and allocation process - it's too complicated for people who are not experts and do not know what their exact needs are. The allocations run out either because we don't know how to use them or we are not using them wisely. There should be a tiered system where light users/beginners have easy access. Help - There is no one to talk to - it would have been so much better for people who might have a lot of questions. Otherwise it is discouraging, takes too long to get through multiples hurdles. Was sent expired links to documentation.
- Restarting machines takes a long time
- I find the relationship between Xsede and Jetstream confusing. I had a lot of trouble navigating the initial signup and login procedures. Globus, single source authentication, Atmosphere, and so-on. The amount of documentation on the Xsede portal is overwhelming. Take a look at Digital Ocean for an example of UI done really well. However, I want to say that I do really appreciate what Jetstream has to offer, and the mission that it fulfilling. What I've described are annoyances that don't diminish the overall excellent value provided by Jetstream.
- Launching images takes too long. Sometimes it's faster to run calculations locally without having to wait for images to launch.
- I found the documentation too hard to understand.
- We are still in the process of integrating our workflows into the environment of Jetstream utilizing the OpenStack API. We only recently had managed to properly connect our services and planned to properly make use of the provided infrastructure of Jetstream,
- Do not find interface easy to navigate and use. Engaging support has felt difficult and sometimes feels like it's difficult to get someone to be engaged.
- It's not necessarily Jetstream's fault, but rather OpenStack's. However, the fact that you provide a bazillion of images makes using Horizon (when it has to load all of them into web browser) unbearable at times. Since I'm using my own images, and only one at a time, it's really annoying.

D.3. Appendix Part 3 — Text Comments for: Please explain how the limitation on the capacity of Jetstream's storage allocations has impacted your ability to perform research.

- Need more storage to process data
- I work with databases, which require different configurations than what is usually available. Adding extra volumes does not allow a database to expand its storage.
- The default storage associated with an VM is somewhat limited. Although I can connect external volume to an VM, Globus doesn't seem to work with external volume.
- Limited storage capacity has made it difficult to index large corpora and datasets
- The disk space allocated with images is too small for doing what the PI I am assisting needs.
- a. some effort to enable my multi-TB allocation 1st time, b. my prior storage of self-purchased multi-TB disks at < \$50/TB is more flexible to needs of data scientists than current Jetstream requirements for equivalent, and much more cost effective than commercial clouds (e.g. Google, AWS) for TBs of science data storage. Offering 5TB to 20 TB data store over several years, plus facile exchange with long-term tape archiving of research data sets seems like an area of improvement for Jetstream that many data-centric scientists would benefit from.
- The indexing files we generate will approximately take up to a few TB of space, but each VM is only having a few hundred GBs. Perhaps we are not using the correct path for large file storage?
- We do have 700.0 GBs storage allocated by the gateway allocation, but we don't know how to connect or mount to our gateway VM.
- I wanted to do data processing on Jetstream, but when I was told that I can get only 500 GB or so I decided to use Stampede2. Later I got a 5 TB which is intended for data dissemination for a science gateway. Below it says it is possible to get much more. That remains to be confusing to me.
- If I could get a machine with, say, 48 cores, 20 TB and PBS or SLURM installed, that would be really useful. Our code is I/O heavy. Would Jetstream be a good solution for that?
- Performance expectations of larger VM images has been a disappointment.
- I think in our case we did not have very much training and so we were unable to use Jetstream to its fullest capacity. In the way that we intended it to use.
- Some of the software we use wants to create a large number of volumes. Each volume is of small size, but the limitation of 10 volumes has been an issue.
- I find that some uses of Jetstream require very little computational power but a lot of I/O, and it would be nice not to burn lots of SUs to get a bit more space. Volumes are very helpful for this but attaching and detaching is not always smooth or straightforward. Even then there are files that would run me out of volume quickly.
- We had to reroute our project twice because of this location, this brought down the whole project.
- We have a terrific allocation on Wrangler but I gather its going away. We are running high-resolution simulations and the output exceed 20 TB so we waste time transferring data on/off instances attached volumes - Globus is fast but not without cost.
- We have to use an externally mounted storage volume to store data that can be accessed by the VMs, which obviously impacts performance.
- I choose to only run small datasets using Jetstream
- Always have to think about the best way to divide the limited storage to among multiple instances.
- It would be useful to have few-core, high-storage VMs so that I can produce extensive data without eating through my allocation faster than necessary.
- Time for transfer to other limited storage
- There have been times when more memory was affordable but we didn't have enough SUs

- The size of the available volumes (20 TB per instance) has made for increased challenges in transferring data to Jetstream for analysis with Matlab. This will become even more challenging in 2020 when we begin WRF simulations on Jetstream.
- The addition of Wrangler mounted to an API instance has greatly assisted our efforts in analysis. This came online for us in mid-2019, and has helped, but we still need larger volumes for our Atmosphere-based instances.
- I used Amazon instead because my allocations were way too small to allow me to store the temp. files (bam files etc.)
- We can't create enough slurm clusters with current quota of 20TB.
- We had large data sets needed by our images that we could not accommodate on Jetstream

D.4. Appendix Part 4 — Text Comments for: Considering your use of Jetstream relative to other national resources (e.g., UCSD, Comet, PSC, Bridges, TACC, Stampede 2, etc.), which of its unique features are most useful to your research program? Other:

- Make it like AWS.
- Ease of installing software packages (much simpler on instances with root access than on shared HPC clusters)
- Ability to create and run customized Docker containers for Jupyter Notebook
- Ability to be root on the custom VMs
- Ability to allow students [to] interact with Jupyter notebooks directly from their browser without need to install any software
- Ability to user Rstudio and install new R packages (e.g., sparklyr for Spark w/R)
- These are all potentially great but I struggled with getting VMs to load without error
- Haven't used any other national resources
- I haven't used other national resources

D.5. Appendix Part 5 — Text Comments for: Please share with us any additional information you feel necessary to understand your storage needs.

- VM can have 1TB storage or easy mount other IU storage system.
- Being able to specify to a person the storage configuration necessary would be better. Connecting to a large disk storage 1-5 TB would be beneficial.
- Most of our storage needs could be met with "cold storage", that is, we have large amounts of data that is accessed infrequently.
- I need a mix of long-term (permanent) storage of published data sets (e.g. 5 TB), including via public archives, multi-year storage of working data sets that are reused, reanalyzed, updated over several years (e.g. 10-15 TB), plus short term/temp storage during analyses (e.g. 5-10 TB for weeks). The working data sets are most useful in an accessible Unix file system amenable to standard Unix search tools (these are bio-data in many file formats, with searchable text).
- Mounts to CVMFS could be useful
- The issue is we use DoE HPSS for permanent storage which is fine AFTER we have completed analyses BUT for the analyses we need access to the full model output (so for time scales of a year or so as while we analyze, write a paper and then respond to review comments)

- I am only using trial version of Jetstream.

D.6. Appendix Part 6 — Text Comments for: For which activities are you using commercial cloud resources?

- 25%
- Custom persistent VMs on AWS, vanilla temporary VMs on Azure, IoT on IBM, API integration with Google (but no compute yet)
- I am not the researcher, the PI is, and he is using both, but the commercial cloud is too expensive
- Commercial clouds are not yet cost effective for research data
- Jetstream lacks support for several cloud services related to Serverless, ML, and streaming data.
- I have used Digital Ocean in past project, currently only using Jetstream
- Backup & Recovery
- For now just S3 storage as it is very easy to use.
- We only use commercial cloud to confirm our systems can function on Jetstream and commercial cloud providers.
- Minimal usage, use AWS for posting free tier community images and occasional testing.
- Some of my collaborators use these services for their work.
- I used Amazon bc I couldn't get Jetstream to work properly.
- I graduate from school, I used Jetstream only in school
- Used AWS for websites. Jetstream for research.
- I do not have enough resources from Jetstream and so I am forced to buy from AWS.

D.7. Appendix Part 7 — Text Comments for: Please share with us any general comments about your experiences with Jetstream. (For example, are there any current features that you find particularly useful? How has the use of Jetstream improved your ability to perform research? Etc.)

- Being a GUI and a being able to do science and not having to be a computer programmer or needing to know Linux is good. Got to get back to the science.
- Very positive, limited storage is a big issue
- Jetstream has been central to gateway hosting and automated post processing of applications logs.
- I found some of the documentation difficult for setting up a persistent IP address. Also, it would be good to have a tutorial around the deployment of scalable resources.
- Currently we're just testing Jetstream to see if it can be used as we want it to be. But I really like the web interface (dashboard) and I use the web shell a lot as well.
- The ability to easily install custom, less widely used software tools thanks to having root access to instances is great--it allows us to experiment and get things going much more quickly than on some of the shared HPC clusters I've used before. And I find that the students tend to be less fearful about making a mistake on a virtual instance than on shared HPC clusters, which in my opinion helps them learn more through trial and error.
- The API is a bit slow, but else it worked remarkably well.
- I wish it had a Kubernetes solution.

- The ability to create VMs and instances quickly and on demand has been very useful when we need to demonstrate our system and in maintaining and running containerized services.
- It would be nice to have more web documentation on how to setup more advanced feature, e.g. a proxy.
- Up-time hours were a little low for a start-up allocation. It was just enough to get through initial development, but not user testing. An extra 10% would have been a real help.
- N/A
- It's allowed me to demonstrate custom Jupyter apps deployed in the cloud (VM + Docker) to other researchers prior to deploying to commercial services. Now looking at Docker Compose and Kubernetes for more complex multi-container workflows.
- Sometimes the server connection is not the most stable. We use Jetstream as servers for running experiments on Amazon Mechanical Turk. Most of the times it works well, but there are times when the connection is lost without any notifications.
- Much of what I use JS for (bioinformatics/genomics) requires a) significant memory b) large core count c) very long or unlimited runtime
- Size of projects continues to increase....JSII might incorporate bigger memory footprint? I currently use Bridges at PSC when hardware limitations are an issue.
- My Ph.D. students use Jetstream for their research and help me with workshop and courses as well. Jetstream helps my students to learn data analytics and cyber security in my class and then continue their education by conducting research leveraging Jetstream resources.
- I hope to use it more as time permits, i.e. as more of a development platform. Now it is (a) a public data server for web-searchable genome data sets (like science gateway service), and (b) intermediate data store for shuffling data among compute servers and home workstations (now that Indiana U has closed off its campus network).
- generally no issues with Jetstream aside from it crashing any time I need to spin up more than half a dozen VM's. For deployments that I grow by hand over a few days, it stays up and running with very few issues for months at a time.
- I love the Jetstream/Atmosphere service, as it allows my faculty at Doane University, a small liberal arts institution, to teach our students about computational strategies and resources in the classroom. With Jetstream, we can have 30 exactly the same learning environments on virtual desktops to level the learning playing field. It really removes the pressure from our students and departments, that generally cannot afford suitable and similar computers to run computational environments, in VMs or otherwise.
- The main problems I have had with Jetstream have been at the TACC data center. During the summer of 2019 we encountered problems with the network availability of some VMs. They would cycle on and off the network which made them unusable. We worked with Jetstream TACC staff to find a resolution, but we were never able to get to the bottom of the problem. It did not seem like the issue was much of a priority for Jetstream TACC. Jetstream at TACC seems like a bit of an afterthought with resources mainly devoted to HPC. I have since shifted all my Jetstream activity onto IU.
- Thank you for your support!
- with traditional HPC (including our campus-based offerings), Jetstream means that my ecology PhD students can quickly dive into large data and computationally complex routines that put their research at the leading edge without facing all the barriers of traditional HPC (shell-based commands, ssh, terminal-based text editors, slurm queues, etc.) and can work with modern tools

they already know and which match commercial offerings (RStudio server, Kubernetes, Docker). Removing this friction opens the door to exploration and experiment that would never be done in the old mode where we develop locally, test thoroughly, and scale up only gradually. I can train scientists instead of being bogged down teaching sys admin skills.

- The Jetstream service has provided critical access to a platform for quickly testing new codes or new workflows. In the educational area, it has allowed students to experience authentic scientific computing on an easy to maintain platform.
- Was used last year for pedagogical purposes
- Jetstream has been great for the Hawaii EPSCoR project in support the diverse needs of our researchers. The staff has been very responsive and helpful in getting us up and running and helping with any issues we encounter. This is a wonderful and indispensable resource for research.
- User management could be improved such that a grant holder can control or observe user VMs.
- I wish to have an easy way to gain access to a Network File System. If there's already one, please put it in somewhere obvious in the documentation.
- I'm extremely grateful for Jetstream. It has definitely improved my ability to perform research by providing me with an accessible way to run hefty processing from almost anywhere.
- Virtual machines not stable during login.
- - I'm running an instance of RStudio in a Docker container that has two different user ids + logins associated with in, so my collaborator and I can work in parallel in the RStudio instance with separate logins
- I like simplicity around spinning up new instances.
- Not sure how to answer issue around storage space as we are using Wrangle and Rancher for these.
- It's helped us build out a service on a tight budget. Decent documentation, responsive customer service team... The only issue has been the extremely slow OpenStack Horizon interface. Thanks y'all!
- Very good. Feels like the system is more stable and highly available than in past years. I prefer it to EC2 instances entirely now.
- make a tutorial of GPU with deep learning (keras or tensorflow)
- the documentation can be more clear
- I like the availability of various necessary tools on VM, since installing them locally sometimes cost a lot of time. Thanks.
- I like to be able to login without MFA or a token.
- I like the potential but I've been struggling with getting images to boot with appropriate sizes. Lots of deployment errors, etc.
- Jetstream has been great for my research. All your hard work has been greatly appreciated.
- Jetstream has made it so much easier to deploy materials to a broad audience. Since we don't need to rely on individual user systems, we can cut out the time spent troubleshooting why things don't work on disparate systems during workshops. The web shell (when it isn't slow/frustrating) can be useful for interactive software and visualizations.
- It appears that the same compute resources (# of cores, same amount of RAM) on Comet, Stampede and Lonestars, and some other private clusters I am using all perform faster and better than the same resource under the VM layer. I don't have hard data to back this up, but the compute performance of Jetstream seems to lag behind the others. Sometimes it appears that the VM layering is causing more problems than it solves. I would prefer an ordinary x86 based cluster without VM on top of it.

- The group of myself , two Post Docs and a graduate student use my Jetstream allocation every single day... We have done research we simply could not have done without Jetstream and that is clearly evidenced by the publications I and my team have produced. Using VNC viewer to mount the instances is brilliant because it makes interacting with Matlabs GUI seamless. The 44 cores (with substantial RAM) with 20 TB mounted volumes means high-resolution numerical simulations and subsequent data analysis are fast and efficient. The biggest challenge for me is I write the proposal to get the resources and am actively involved in the research but a Post Doc manages the instances - so I functionally have ZERO access to information about the instances in case of problems.
- It would be nice to transfer VM's between OpenStack API and Atmosphere
- Jetstream has been absolutely amazing for hosting our science gateway. The Jetstream team is amazing and has helped resolve every issue we have had. This unique capability was a great investment - thanks Jetstream and NSF!
- It's absolutely amazing that such a resource should be available and we would not be able to do our research without it. The only real issues have been (1) getting up and running and (2), possibly relatedly, the web interface. (For example, we did not even immediately know we had been approved for an allocation)
- Custom VMs with bioinformatic tools are awesome!
- I've been extremely pleased with the ability to reach the Jetstream/XSEDE computing resources from anywhere with an internet connection. I would highly recommend to anyone else interested in accessing this capability.
- Experience is overall positive. For whatever reason, we had issue with accessing instances (ssh timeout; took months to resolve). But not having to wait in a queue is very useful.
- We would benefit from a more detailed resource-management of an allocation from the XSEDE page.
- We are using Jetstream primarily for education purposes by hosting web servers and gateways. Jetstream has provided the ability to integrate with our local HPC resources for computing, as well as connecting to other national cloud platforms to provide service to users. This is very useful and we would like to keep it going for longer periods of time in future.
- Availability of Jetstream has enabled me to partake in computational research previously not possible. I have been able to compute on large datasets and be involved in additional projects that required computational allocations not available to the collaborators. The flexibility of resource choice along with speedy issue resolution makes it one of the best platforms I have performed computation.
- It is essential to have access for high-resolution numerical modeling. There is no other solution.
- I have shown demos of our services (online calculations) with seamless performance, every time.
- While I like the idea of having access to computing resources through Jetstream's VMs I was not able to work with them. Much of my allocations run out without performing any true computation e.g., just figuring out how to move files from a server to Jetstream etc.
- Atmosphere is a good tool for teaching workshops/classes
- A better description of available images.
- Jetstream has become an essential component of our NSF sponsored research. We develop a tool to build science gateways, so it is helpful to test correctness on various operating systems. It has been *extremely* useful to quickly setup VMs for postdocs, graduate, undergraduate and high school students and summer intern development projects so that they can have their own "sandboxes" to which we also have access. We also use Jetstream heavily to host project pages and multiple science

gateways which are used in training and production. API usage is good for persistent instances and elastic computing. We tend to prefer the ease of Atmosphere to quickly get a VM running.

- This past year I did not get much opportunity to use Jetstream for reasons other than to do with Jetstream issues.
- I don't quite understand what it is and how it can be helpful for my team and me. I would mostly prefer one on one consultations for "training" when I'm getting started.
- It is time to bring down the silos and make room for collaboration, if only read only
- Jetstream is an essential part of my research program.
- As the Assistant Director of Research Computing for the University of Arizona I really appreciate using Jetstream to develop and deploy CI code as well as send researchers to Jetstream to do their work. Availability of on-demand computing for researchers is exceedingly rare, and is crucial for developing analysis pipelines, research cyberinfrastructure code, as well as building and testing containers. Jetstream has made our Software and Data Carpentry trainings easier as participants can learn how to access remote resources, have the same computing environments to minimize computational variation problems, and don't have to worry about 'hurting anything on their computer.' That is to say: if they delete system files, break package/library installs, or anything else they can just restart a new VM with a clean install. Thank you for making my life easier in all regards.
- My team could not conduct the research we do without Jetstream. The availability of flexible computational resources is invaluable for our lab. Jetstream is also a fabulous resource for introducing novices to HPC and cloud computing.
- The flexibility offered by the Openstack API has been a great benefit.
- I like the great training classes and various images of Jetstream.
- We are having a very positive experience. After a very difficult interaction with AWS and a lot of issues when using AWS allocations in class, we moved to Jetstream and since then we do not miss AWS at all. The students are very pleased and we are able to move forward quickly with the training and teaching. Jetstream is now a key part of our work in class. My students are also using Jetstream in their daily research successfully. Thank you!
- Jetstream is good to do experiments on.
- Bringing data in and out of the environment is pretty clunky. The atmosphere environment is not something I would code in. I'd probably code locally and push code to the VM.
- The support team has been very responsive and effective. Thanks for your help and generosity!
- Access to dedicated cloud resources has reduced time to science, and increased research output (publications). We are still bound by RAM constraints. This has been an issue for analysis larger datasets. We would really benefit from more than 120 GB per instance.
- For us to be able to continue using Jetstream we need GPUs, and not for ML or computing but for interactive 3D rendering and visualization
- Jetstream has particularly been invaluable for running bioinformatics workshops, enabling the handling and configuration of virtual machines for over a hundred participants at a time. Thank you for all you do, folks :)
- Cumbersome to set up, not enough allocation to do RNA seq
- I think it's good, sometimes has bugs, but overall good resources
- The help from staff is incredibly useful and rapid. I have had difficulty finding appropriate help articles online to onboard new users. Also, I feel like there are probably services that would help us

for which we are not familiar (e.g., we don't typically use Jetstream storage, but wonder if it would be helpful).

- I attended a two-week workshop and training in data science where I used Jetstream first time. Most important of Jetstream is that you need to bother about OS and tool you can simply use an image.
- Keep up the great work – Jetstream has been a huge help for our research group over the past 5 years.
- Jetstream allows us to deploy our development, test, and production instances of our service stack. It's been fundamental for our platform and it also allows us to run slurm clusters which without long queue time which is important to accelerate scientific discovery made through our platform.
- Jetstream is truly unique and invaluable service that I could not complete my research without. In addition, we use Jetstream to run a Science Gateway, and that empowers researchers across the country to answer heretofore unanswerable questions. We could not do research without Jetstream
- Jetstream is vital to my research. We configure heterogeneous set of VMs and conduct research on how container orchestration platforms can be used in cloud settings. My request is to increase the allocation in Jetstream so that we can also conduct scalability studies.
- Thanks for all your great work on behalf of the community! Kudos to Jeremy who is my primary interaction with Jetstream support -- he's amazing.
- availability of cluster resources is very helpful, but nodes are occasionally unavailable for unknown reasons.
- Due to difficulties of setting up the API and integration thereof with our internal systems, we weren't able to evaluate Jetstream sufficiently within the given timeframe of the allocation.
- Without Jetstream we would be unable to provide the services we currently offer to the scientific community.

D.8. Appendix Part 8 — Text Comments for: Please provide citations for the products that you indicated have benefitted from the use of Jetstream, including title, authors, publication, publication date, when and where presented, URL, and/or digital object identifier (DOI).

- Papudeshi, B., Sanders, S., Ganote, C., Doak, T. (2019). Compute resources available to the research community for microbiome analysis. In Plant and Animal Genomics 2019, San Diego, CA. Retrieved from <http://hdl.handle.net/2022/22655>
- Papudeshi, B., Sanders, S., Ganote, C., Doak, T., Chafin, T., Reshetnikov, A., Sokolov, S., Pummil, J., Douglas, M., Douglas, M. (2019). The Genome of Fish Tapeworm *Nippotaenia percotti* as a Potential Bookmark for Gene Loci that Facilitates Anthropogenic Infection. In Plant and Animal Genomics 2019, San Diego, CA. Retrieved from <http://hdl.handle.net/2022/22656>
- Cai, Jasmine X.; Weathers, Jania G.; Leffler, Haley; Ganapaneni, Sruthi; Papudeshi, Bhavya; Sanders, Sheri; Doak, T. G. (2019) Navigating the Sequence Read Archive to identify crAssphage, a ubiquitous inhabitant of the human microbiome, in. Bloomington, Indiana: Jim Holland Summer Science Research Program Poster Session.
- Leffler, H. et al. (2019) A workflow to identify genomes in the Sequence Read Archive for phylogenomic analysis, in. San Francisco: ASM Microbe 2019. <http://hdl.handle.net/2022/23371>

- Leffler, H. et al. (2019) Mining the Sequence Read Archive to identify crAssphage, a ubiquitous inhabitant of the human microbiome, in dog and pig samples.
<https://scholarworks.iu.edu/dspace/handle/2022/23210>
- Ganapaneni, S. et al. (2019) Coupling metagenomics with high-performance computing to mine the Sequence Read Archive (SRA) to analyze Pseudomonas phage PAK-P1.
<https://scholarworks.iu.edu/dspace/handle/2022/23209>
- Donnellan, A., Parker, J., Heflin, M., Lyzenga, G., Moore, A., Ludwig, L. G., et al. (2018). Fracture advancing step tectonics observed in the Yuha Desert and Ocotillo, CA, following the 2010 Mw7.2 El Mayor, Cucapah earthquake. *Earth and Space Science*, 5, 456, 472.
<https://doi.org/10.1029/2017EA000351>
- gridchem.uits.iu.edu
- <http://gridchem.uits.iu.edu:8080/axis2/>
- https://gridchem.uits.iu.edu/open_inventory_2019-04-24/
- <https://django.seagrid.org/>
- Yinzhi Wang, Sudhakar Pamidighantam, Sean Yaw, Eroma Abeysinghe, Suresh Marru, Marcus Christie, Kevin Ellett, Marlon Pierce, and Richard Middleton. 2018. A New Science Gateway to Provide Decision Support on Carbon Capture and Storage Technologies. In *Proceedings of the Practice and Experience on Advanced Research Computing (PEARC '18)*. ACM, New York, NY, USA, Article 76, 3 pages. DOI: <https://doi.org/10.1145/3219104.3229244>
- Marlon Pierce, Suresh Marru, Eroma Abeysinghe, Sudhakar Pamidighantam, Marcus Christie, and Dimuthu Wannipurage. 2018. Supporting Science Gateways Using Apache Airavata and SciGaP Services. In *Proceedings of the Practice and Experience on Advanced Research Computing (PEARC '18)*. ACM, New York, NY, USA, Article 99, 4 pages. DOI: <https://doi.org/10.1145/3219104.3229240>
- Science Gateway Use Cases, version 1.1
- Marru, Suresh; Alameda, Jay; Gunathilake, Lahiru; Liu, Yan; Martin, Stuart; Middleton, Don; Palencia, Josephine; Pamidighantam, Sudhakar; Pierce, Marlon; Schwartz, Terri; Singh, Raminder; Thompson, Chris; Uram, Tom; Wang, Shaowen; Wilkins, Nancy; Xu, Haiying;
<http://hdl.handle.net/2142/104719>
- Harsha Phulwani, Madrina Thapa, Suresh Marru, Marlon Pierce, Sudhakar Pamidighantam, and Marcus Christie. 2018. Apache Airavata Resource Allocation System: A Tool for Allocating Resources in Science Gateways. In *Proceedings of the Practice and Experience on Advanced Research Computing (PEARC '18)*. ACM, New York, NY, USA, Article 78, 4 pages. DOI: <https://doi.org/10.1145/3219104.3229271>
- Towards Run Time Estimation of the Gaussian Chemistry Code for SEAGrid Science Gateway
Angel Beltre, Shehtab Zaman, Kenneth Chiu, Sudhakar Pamidighantam, Xingye Qiao, Madhusudhan Govindaraju; [10.1145/3332186.3338101](https://doi.org/10.1145/3332186.3338101)
- Adison A. Kleinsasser, Sudhakar Pamidighantam, Douglas M. Jennewein, Joseph D. Madison, Marcus Christie, Eroma Abeysinghe, Suresh Marru, and Marlon Pierce. 2019. The USD Science Gateway: A Bridge Between Research and Advanced Computing. In *Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines (learning) (PEARC '19)*. ACM, New York, NY, USA, Article 125, 4 pages. DOI: <https://doi.org/10.1145/3332186.3333254>
- Joseph D. Madison, Eroma Abeysinghe, Sudhakar Pamidighantam, Suresh Marru, Marcus Christie, Douglas M. Jennewein, and Marlon Pierce. 2018. Science Gateway Implementation at the University of South Dakota: Applications in Research and Education. In *Proceedings of the Practice and*

Experience on Advanced Research Computing (PEARC '18). ACM, New York, NY, USA, Article 97, 4 pages. DOI: <https://doi.org/10.1145/3219104.3229265>

- Dimuthu Wannipurage, Suresh Marru, Marlon Piece, Eroma Abeysinghe, Sudhakar Pamidighantam, Marcus Christie, Gourav Shenoy, Ajinkya Dhamnaskar, and Lahiru Jayathilaka. 2019. Implementing a Flexible, Fault Tolerant Job Management System for Science Gateways. In Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines (learning) (PEARC '19). ACM, New York, NY, USA, Article 15, 8 pages. DOI: <https://doi.org/10.1145/3332186.3332233>
- Eric Lease Morgan, Eroma Abeysinghe, Sudhakar Pamidighantam, Eric Coulter, Suresh Marru, and Marlon Pierce. 2019. The Distant Reader: Tool for Reading. In Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines (learning) (PEARC '19). ACM, New York, NY, USA, Article 124, 4 pages. DOI: <https://doi.org/10.1145/3332186.3333260>
- American Astronomical Society, AAS Meeting #233, id.242.2
- When we get our application (<https://vamps2.mbl.edu>) properly functioning on Jetstream we'll have citation data for various projects for example:
<http://codl.coas.oregonstate.edu/>
<http://icomm.mbl.edu/>
But not yet :)
- Baughman et al. (2019) "Deconstructing the 2017 Changes to AWS Spot Market Pricing." 10th Workshop on Scientific Cloud Computing, in proceedings. DOI: 10.1145/3322795.3331465
- Knepper, R., et al. (2017). Using the Jetstream Research Cloud to provide Science Gateway resources. Proceedings of the 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, IEEE Press.
- Perri, M. J. (2018). Chem Compute Educational Gateway. SGCI - MolSSI American Chemical Society Workshop on Science Gateways, New Orleans, LA.
- Perri, M. J., et al. (2019). Chem Compute Science Gateway: An Online Computational Chemistry Tool. Using Computational Methods To Teach Chemical Principles, American Chemical Society. 1312: 79-92.
- Perri, M. J., et al. (2019 (In Press)). Chem Compute Science Gateway: An Online Computational Chemistry Tool. Using Computational Methods to Teach Chemical Principles. A. Grushow and M. Reeves, ACS Press.
- Perri, M. J., et al. (2019). Chem Compute Science Gateway for Undergraduates. American Chemical Society National Meeting, Orlando, FL.
- Perri, M. J., et al. (2017). Chem Compute science gateway: Web-based computational job submission for the undergraduate laboratory. American Chemical Society National Meeting, Washington, DC.
- Reeves, M., et al. (2017). How can you measure a reaction enthalpy without going into the lab?: Using computational chemistry data to draw a conclusion. American Chemical Society National Meeting, Washington, DC.
- Reeves, M., et al. (2019 (In Press)). How can you measure a reaction enthalpy without going into the lab? Using computational chemistry data to draw a conclusion. Using Computational Methods to Teach Chemical Principles. A. Grushow and M. Reeves, ACS Press.
- Reeves, M. S., et al. (2019). How Can You Measure a Reaction Enthalpy without Going into the Lab? Using Computational Methods To Teach Chemical Principles, American Chemical Society. 1312: 51-63.

- Thiemann, N. (2017). Using Computational Tools to Design a Molecularly Imprinted Polymer with Selectivity and High Affinity for Acetylcholine. Neuroscience Program. Hartford, Connecticut, Trinity College. Baccalaureate Degree with Honors in Neuroscience: 29.
- Wiens, J. P. and W. A. Alexander (2019). "Sodium Atom Beam Collisions with the Liquid Glycerol Surface: Mass Effects of Deuteration." Chemical Physics Letters.
- R. Eckart de Castilho, N. Ide, J. Kim, J. Klie, K. Suderman. A Multi-Platform Annotation Ecosystem for Domain Adaptation. Proceedings of the 13th Linguistic Annotation Workshop. 189--194. <https://www.aclweb.org/anthology/W19-4021>. 2019.
- R. Eckart de Castilho, N. Ide, J. Kim, J. Klie, K. Suderman. Towards cross-platform interoperability for machine-assisted text annotation. Genomics & Informatics. 2. e19. <http://dx.doi.org/10.5808/GI.2019.17.2.e19>. 2019.
- Erhard Hinrichs, Nancy Ide, James Pustejovsky, Jan Hajic, Marie Hinrichs, Mohammad Fazleh Elahi, Keith Suderman, Marc Verhagen, Kyeongmin Rim, Pavel Stranak, Josef Misutka. Bridging the LAPPS Grid and CLARIN. Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). Miyazaki, Japan. 2018.
- Nancy Ide, J. Kim, K. Suderman. Mining Biomedical Publications With The LAPPS Grid. Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). Miyazaki, Japan. 2018.
- Deploying Jupyter Notebooks at scale on XSEDE resources for Science Gateways and workshops, A Zonca, R Sinkovits, PEARC18, <https://arxiv.org/abs/1805.04781>
- 10.1145/3332186.3333260
- Ioannis Paraskevatos, Andre Luckow, Mahzad Khoshlessan, George Chantzialexiou, Thomas E. Cheatham, Oliver Beckstein, Geoffrey C. Fox, and Shantenu Jha. 2018. Task-parallel Analysis of Molecular Dynamics Trajectories. In Proceedings of the 47th International Conference on Parallel Processing (ICPP 2018). ACM, New York, NY, USA, Article 49, 10 pages. DOI: <https://doi.org/10.1145/3225058.3225128>
- Ioannis Paraskevatos, Matteo Turrili, Bento Collares Gonçalves, Heather J. Lynch and Shantenu Jha, "Workflow Design Analysis for High Resolution Satellite Image Analysis", 2019, <http://arxiv.org/abs/1905.09766>
- N/A
- Gilbert, DG. 2019. Genes of the Pig, *Sus scrofa*, reconstructed with EvidentialGene. PeerJ 7:e6374; doi:10.7717/peerj.6374
- Lindeman et al 2019. doi: 10.1093/eep/dvz016 (use public data from eugenes.org on Jetstream) many others use software, documents from eugenes.org
- ~30 pubs in 2019, see in <http://arthropods.eugenes.org/EvidentialGene/evigene/docs/evigene-cites.txt>
- Not really applicable, but I wanted you guys to know.
- http://science-gateway.unidata.ucar.edu/jetstream_bib.html
- Ramamurthy, M. K. and Chastang, J. and May, R. M. and James, M., Unidata and data-proximate analysis and visualization in the cloud. ESIP Summer Meeting 2017 <https://doi.org/10.6084/m9.figshare.5249839.v1>
- An overview of the {XSEDE} extended collaborative support program. Nancy Wilkins-Diehr and Sergiu Sanielevici and Jay Alameda and John Cazes and Lonnie Crosby and Marlon Pierce and Ralph

Roskies, 2016. 10.1007/978-3-319-32243-8_1 High Performance Computer Applications - 6th International Conference, ISUM 2015, Revised Selected Papers.

- Chastang, J. and Yoksas, T. and Ramamurthy, M. K., Geoscientific Data Distribution in the {XSEDE} {Jetstream} Cloud. Proceedings, 33rd Conference on Environmental Information Processing Technologies, 97th AMS Annual Meeting.
<https://ams.confex.com/ams/97Annual/webprogram/Paper315508.html>
- Chastang, J. and Signell, R. Met/Ocean Modeling Workflows on {XSEDE} via {HPC} \& Cloud}, ESIP Summer Meeting 2017, 10.6084/m9.figshare.5249845.v1.
- Wilcox, K. and Chastang, J. Unidata THREDDS Docker Container, 2015.
<https://github.com/Unidata/thredds-docker>
- Unidata LDM Docker Container. Chastang, J. GitHub repository, <https://github.com/Unidata/lmd-docker/>
- Unidata on the XSEDE {Jetstream} Cloud. Chastang, J., 2017, <https://github.com/Unidata/xsede-jetstream>.
- Chastang, J. and Ramamurthy, M. K. Unidata Science Gateway on the XSEDE {Jetstream} Cloud, 10.6084/m9.figshare.5479648.v1
- Chastang, J. and Signell, R. and Fischer, J. L. Reducing Time to Science: Unidata and {JupyterHub} Technology Using the Jetstream Cloud, 2017, <https://doi.org/10.6084/m9.figshare.5660128.v1>
- Chastang, J. and Signell, R. and Fischer, J. L. A {Unidata} {JupyterHub} Server: An Online {PyAOS} Resource for Students and Educators, <https://doi.org/10.6084/m9.figshare.5756643.v1>,
- Ramamurthy, M. K. and Chastang, J., A Cloud-based Science Gateway for the Geoscience Community with End-to-end Workflows on the Jetstream Cloud System, See also <http://science-gateway.unidata.ucar.edu>
- Chastang, J. and Ramamurthy, M. K. A Cloud-based Science Gateway for the Geoscience Community, <https://doi.org/10.6084/m9.figshare.6451508.v1>
- Sarajlic, S. and Chastang, J. and Marru, S. and Fischer, J. and Lowe, M., Scaling JupyterHub using Kubernetes on Jetstream Cloud: Platform as a Service for research and educational initiatives in the Atmospheric Sciences, See also <http://science-gateway.unidata.ucar.edu>
- Ramamurthy, M. K. and Chastang, J. A Cloud-based Science Gateway for the Geoscience Community with End-to-end Workflows on the {Jetstream} Cloud System, See also <http://science-gateway.unidata.ucar.edu>
- Ramamurthy, M. K. A Cloud-based Science Gateway for the Geoscience Community, <https://doi.org/10.6084/m9.figshare.6856733.v1>
- Chastang, J. Unidata Science Gateway JupyterHub. <https://doi.org/10.6084/m9.figshare.6856733.v1>
- Ramamurthy, M. K. and Chastang, J. Unidata: A Cloud-based Science Gateway for the Atmospheric Sciences, See also <http://science-gateway.unidata.ucar.edu>
- Chastang, J. JupyterHub for Atmospheric Science Research and Education on the {NSF} {J}etstream Cloud. <https://doi.org/10.6084/m9.figshare.7590869.v1>
- Ramamurthy, M. K. and Chastang, J. A Cloud-Based Science Gateway for Enabling Open and Reproducible Science, <https://ams.confex.com/ams/2019Annual/meetingapp.cgi/Paper/354804>
- Chastang, J. Deploying a {U}nidata {J}upyter{H}ub on the {NSF} {J}etstream Cloud, Lessons Learned and Challenges Going Forward, <https://doi.org/10.6084/m9.figshare.8945078.v1>
- doi:10.1073/pnas.1902657116

- doi:10.1086/702704
- doi:10.21105/joss.01276
- doi:10.21105/joss.00971
- doi:10.1016/j.biocon.2018.05.009
- doi:10.5281/zenodo.1486394
- doi:10.5281/zenodo.2528507
- <http://g2s3.com/labs/>
- https://github.com/uvilla/cmisis_labs/
- <https://sinews.siam.org/Details-Page/students-tackle-bayesian-inverse-problems-in-the-colorado-rockies>
- Sutlief, Arin L., Helena Valquier-Flynn, Christina Wilson, Marco Perez, Hunter Kleinschmidt, Brett J. Schofield, Elizabeth Delmain, Andrea E. Holmes, and Christopher D. Wentworth. „Live Cell Analysis of Shear Stress on Pseudomonas Aeruginosa Using an Automated Higher-Throughput Microfluidic System. *JoVE (Journal of Visualized Experiments)*, no. 143 (January 16, 2019): e58926. <https://doi.org/10.3791/58926>.
- Vaughn, S. and Wentworth, C.D. Fourier Analysis of Microscopic Images of Pseudomonas aeruginosa Biofilms Grown Under Changing Shear Stress Conditions Nebraska Academy of Sciences 2019 Annual Meeting Nebraska Wesleyan Lincoln, Nebraska April 12, 2019
- I'm not sure how to fill this in, since the project is still ongoing, so I haven't specifically presented the data that we use Jetstream for yet (analysis ongoing), but have presented about the project in general at two industry meeting, a national lab workshop, and a national lab symposium. I will definitely cite Jetstream when presenting the data analyzed on it.
 - Presentations:
 - Center for the Built Environment industry meeting, Berkeley CA, April 25 2019
 - Center for the Built Environment industry meeting, October 2018
 - "Cool Buildings" workshop, Lawrence Berkeley National Lab, July 22 2019
 - Rosenfeld Symposium, Berkeley CA, April 23, 2019
- Susa Oram, Cecile Kenny, Eli Sobel, Jeremy Coate Genetic dissection of salt tolerance in Arabidopsis polyploids. Botany 2019 (Tucson, AZ). Manuscript not yet submitted.
- We're using Jetstream to build out StochSS 2, a total rebuild of the Stochastic Simulation Service focusing on a scalable service infrastructure. <https://github.com/StochSS/stochss/tree/develop>
- <https://ieeexplore.ieee.org/abstract/document/8588659>
- Wang, Zhong, Marcus A. Christie, Eroma Abeysinghe, Tinyi Chu, Suresh Marru, Marlon Pierce, and Charles G. Danko. "Building a science gateway for processing and modeling sequencing data via Apache Airavata." In Proceedings of the Practice and Experience on Advanced Research Computing, p. 39. ACM, 2018.
- <https://doi.org/10.1111/1462-2920.14552>
- <https://doi.org/10.1093/bioinformatics/btz188>
- N/A
- Single and population coding of taste in the gustatory cortex of awake mice. Levitan D, Lin JY, Wachutka J, Mukherjee N, Nelson SB, Katz DB. *J Neurophysiol.* 2019 Oct 1;122(4):1342-1356. doi: 10.1152/jn.00357.2019. Epub 2019 Jul 24.

- Avesani, P., Caiafa, C., McPherson, B. Saykin, A., Hayashi, S., Herschel, R., A., Garyfallidis, E., Kitchell, L., Bullock, D., Patterson, A., O'Riley, S., Olivetti, E., Sporns, O., Saykin, A., Wang, L., Dinov, I., and Pestilli, F. (2019) The open diffusion data derivatives, brain data upcycling via integrated publishing of derivatives and reproducible open cloud services. *Nature: Scientific Data*. DOI: 10.1038/s41597-019-0073-y.
- Demeler, B. Measuring molecular interactions in solution using multi-wavelength analytical ultracentrifugation: combining spectral analysis with hydrodynamics. *Biophysics - using physics to explore biological systems The Biochemist* 41(2). 2019. 14-18.
- Hu J, Soraiz EH, Johnson CN, Demeler B, Brancalion L. Novel combinations of experimental and computational analysis tested on the binding of metalloprotoporphyrins to albumin. *Int J Biol Macromol*. 2019 May 9. pii: S0141-8130(19)31864-1. doi: 10.1016/j.ijbiomac.2019.05.060.
- Ranaivoson FM, Turk LS, Ozgul S, Kakehi S, von Daake S, Lopez N, Trobiani L, De Jaco A, Denissova N, Demeler B, Añzkan E, Montelione GT, Comoletti D.A Proteomic Screen of Neuronal Cell-Surface Molecules Reveals IgLONs as Structurally Conserved Interaction Modules at the Synapse. *Structure*. 2019 Mar 23. pii: S0969-2126(19)30082-6. doi: 10.1016/j.str.2019.03.004.
- Bansal A, Karanth NM, Demeler B, Schindelin H, Sarma SP. Crystallographic Structures of IlvN- Σ Val/Ile Complexes: Conformational Selectivity for Feedback Inhibition of Aceto Hydroxy Acid Synthases. *Biochemistry*. 2019 Apr 16;58(15):1992-2008. doi: 10.1021/acs.biochem.9b00050.
- Wang Z, Bhattacharya A, White T, Buffone C, McCabe A, Nguyen LA, Shepard CN, Pardo S, Kim B, Weintraub ST, Demeler B, Diaz-Griffero F, Ivanov DN. Functionality of Redox-Active Cysteines Is Required for Restriction of Retroviral Replication by SAMHD1. *Cell Rep*. 2018 Jul 24;24(4):815-823. doi: 10.1016/j.celrep.2018.06.090. PMID: 30044979
- Salveson PJ, Haerianardakani S, Thuy-Boun A, Yoo S, Kreutzer AG, Demeler B, Nowick JS. Repurposing Triphenylmethane Dyes to Bind to Trimers Derived from ACE \leq . *J Am Chem Soc*. 2018 Sep 19;140(37):11745-11754. doi: 10.1021/jacs.8b06568. Epub 2018 Sep 6. Erratum in: *J Am Chem Soc*. 2018 Nov 14;140(45):15546. PMID: 30125493
- Salveson PJ, Haerianardakani S, Thuy-Boun A, Yoo S, Kreutzer AG, Demeler B, Nowick JS. Correction to "Repurposing Triphenylmethane Dyes To Bind to Trimers Derived from ACE \leq ". *J Am Chem Soc*. 2018 Nov 14;140(45):15546. doi: 10.1021/jacs.8b11195. Epub 2018 Nov 5. PMID: 30394740
- Sterritt OW, Lang EJM, Kessans SA, Ryan TM, Demeler B, Jameson GB, Parker EJ. Structural and functional characterization of the entry point to pyocyanin biosynthesis in *Pseudomonas aeruginosa* defines a new 3-deoxy-d-arabino-heptulosonate 7-phosphate synthase subclass. *Biosci Rep*. 2018 Oct 17;38(5). pii: BSR20181605. doi: 10.1042/BSR20181605. Print 2018 Oct 31. PMID: 30242059
- English LR, Tischer A, Demeler AK, Demeler B, Whitten ST. Sequence Reversal Prevents Chain Collapse and Yields Heat-Sensitive Intrinsic Disorder. *Biophys J*. 2018 Jul 17;115(2):328-340. doi: 10.1016/j.bpj.2018.06.006. PMID: 30021108
- Kim H, Brookes E, Cao W, Demeler B. Two-dimensional grid optimization for sedimentation velocity analysis in the analytical ultracentrifuge. *Eur Biophys J*. 2018 Oct;47(7):837-844. doi: 10.1007/s00249-018-1309-z. PMID: 29777290
- Williams TL, Gorbet GE, Demeler B. Multi-speed sedimentation velocity simulations with UltraScan-III. *Eur Biophys J*. 2018 Oct;47(7):815-823. doi: 10.1007/s00249-018-1308-0. PMID: 29748855
- Johnson CN, Gorbet GE, Ramsower H, Urquidi J, Brancalion L, Demeler B. Multi-wavelength analytical ultracentrifugation of human serum albumin complexed with porphyrin. *Eur Biophys J*. 2018 Jul 17;115(2):328-340. doi: 10.1007/s00249-018-1301-7.

- Lopez P, Lara HH, Mullins SM, Black DM, Ramsower HH, Alvarez MM, Williams TL, Lopez-Lozano X, Weissker HC, Garcia AP, Garzan IL, Demeler B, Lopez-Ribot JL, Yacamán MJ, and Whetten RL. Tetrahedral (T) Closed-Shell Cluster of 29 Silver Atoms & 12 Lipolate Ligands, [Ag₂₉(R-CE±-LA)₁₂](3-): Antibacterial and Antifungal Activity. *ACS Appl. Nano Mater.*, 2018, 1 (4), pp 1595-1602
- Gorbet GE, Mohapatra S, Demeler B. Multi-speed sedimentation velocity implementation in UltraScan-III. *Eur Biophys J.* 2018 Oct;47(7):825-835. doi: 10.1007/s00249-018-1297-z.
- Serwer P, Wright ET, Demeler B, Jiang W. States of phage T3/T7 capsids: buoyant density centrifugation and cryo-EM. *Biophys Rev.* 2018 Apr;10(2):583-596. doi: 10.1007/s12551-017-0372-5. Epub 2017 Dec 14. Review. PMID: 29243090
- Abbott JA, Meyer-Schuman R, Lupo V, Feely S, Mademan I, Oprescu SN, Griffin LB, Alberti MA, Casanovas C, Aharoni S, Basel-Vanagaite L, Zochner S, De Jonghe P, Baets J, Shy ME, Espinós C, Demeler B, Antonellis A, Francklyn C. Substrate interaction defects in histidyl-tRNA synthetase linked to dominant axonal peripheral neuropathy. *Hum Mutat.* 2018 Mar;39(3):415-432. doi: 10.1002/humu.23380. Epub 2017 Dec 26. PMID: 29235198
- Wagner JM, Christensen DE, Bhattacharya A, Dawidziak DM, Roganowicz MD, Wan Y, Pumroy RA, Demeler B, Ivanov DN, Ganser-Pornillos BK, Sundquist WI, Pornillos O. A general model for retroviral capsid pattern recognition by TRIM5 proteins. *J Virol.* 2018 Jan 30;92(4). pii: e01563-17. doi: 10.1128/JVI.01563-17. Print 2018 Feb 15. PMID: 29187540
- Tian S, Yu G, He H, Zhao Y, Liu P, Marshall AG, Demeler B, Stagg SM, Li H. Pih1p-Tah1p Puts a Lid on Hexameric AAA+ ATPases Rvb1/2p. *Structure.* 2017 Oct 3;25(10):1519-1529.e4. doi: 10.1016/j.str.2017.08.002. Epub 2017 Sep 14. PMID: 28919439
- Abbott JA, Guth E, Kim C, Regan C, Siu VM, Rupar CA, Demeler B, Francklyn CS, Robey-Bond SM. The Usher Syndrome Type IIIB Histidyl-tRNA Synthetase Mutation Confers Temperature Sensitivity. *Biochemistry.* 2017 Jul 18;56(28):3619-3631. doi: 10.1021/acs.biochem.7b00114. Epub 2017 Jul 7. PMID: 28632987
- Fan Y, Guo YR, Yuan W, Zhou Y, Holt MV, Wang T, Demeler B, Young NL, Zhong W, Tao YJ. Structure of a pentameric virion-associated fiber with a potential role in Orsay virus entry to host cells. *PLoS Pathog.* 2017 Feb 27;13(2):e1006231. doi: 10.1371/journal.ppat.1006231. eCollection 2017 Feb. PMID: 28241071
- Kim SK, Barron L, Hinck CS, Petrunak EM, Cano KE, Thangirala A, Iskra B, Brothers M, Vonberg M, Leal B, Richter B, Kodali R, Taylor AB, Du S, Barnes CO, Sulea T, Calero G, Hart PJ, Hart MJ, Demeler B, Hinck AP. An Engineered TGF- β Monomer that Functions as a Dominant Negative to Block TGF- β Signaling. *J Biol Chem.* 2017 Apr 28;292(17):7173-7188. doi: 10.1074/jbc.M116.768754. Epub 2017 Feb 22. PMID: 28228478
- Zhang J, Pearson JZ, Gorbet GE, Cölfen H, Germann MW, Brinton MA, Demeler B. Spectral and Hydrodynamic Analysis of West Nile Virus RNA-Protein Interactions by Multiwavelength Sedimentation Velocity in the Analytical Ultracentrifuge. *Anal Chem.* 2017 Jan 3;89(1):862-870. PMID: 27977168
- Walter J, GE Gorbet, T Akdas, D Segets, B Demeler, W. Peukert. 2D analysis of polydisperse core-shell nanoparticles using analytical ultracentrifugation *Analyst*, 2016, 142(1):206-217. PMID: 27934989
- Ariella L Gladstein, Michael F Hammer, Substructured Population Growth in the Ashkenazi Jews Inferred with Approximate Bayesian Computation, *Molecular Biology and Evolution*, Volume 36, Issue 6, June 2019, Pages 1162-1171, <https://doi.org/10.1093/molbev/msz047>

- Gladstein, A.L., Quinto-Cortés, C.D., Pistorius, J.L., Christy, D., Gantner, L., Woerner, A.E., Joyce, B.L. (2018). SimPrily: A Python framework to simplify high-throughput genomic simulations. *SoftwareX*, 7:335-340. <https://doi.org/10.1016/j.softx.2018.09.003>.
- Pryor S.C., Shepherd T.J., Volker P., Hahmann A.N. and Barthelmie R.J.: „Wind theft,“ from onshore wind turbine arrays: Sensitivity to wind farm parameterization and resolution. (in review) Letson F.W., Barthelmie R.J., and Pryor S.C.: RADAR-derived precipitation climatology for wind turbine blade leading edge erosion. *Wind Energy Science Discussions* (in review, <https://doi.org/10.5194/wes-2019-43>).
- Shepherd T.J., Barthelmie R.J. and Pryor S.C. Sensitivity of wind turbine array downstream effects to the parameterization used in WRF. (in review)
- Pryor S.C., Shepherd T.J., Barthelmie R.J., Hahmann A., Volker P. (2019): Wind farm wakes simulated using WRF. *Journal of Physics: Conference Series* 1256 012025 doi: 10.1088/1742-6596/1256/1/012025.
- Pryor S.C., Shepherd T.J. and Barthelmie R.J. (2018): Inter-annual variability of wind climates and wind turbine annual energy production. *Wind Energy Science* 3 651-665.
- Pryor S.C., Barthelmie R.J. and Shepherd T.J. (2018): The influence of real-world wind turbine deployments on local to mesoscale climate. *Journal of Geophysical Research: Atmospheres* 123 5804-5826 (doi: 10.1029/2017JD028114).
- Pryor S.C., Barthelmie R.J., Hahmann A., Shepherd T.J., Volker P. (2018): Downstream effects from contemporary wind turbine deployments. *Journal of Physics: Conference Series* 1037 072010 doi :10.1088/1742-6596/1037/7/072010.
- Pryor S.C. and Hahmann A.N. (2019): Downscaling wind. In *Oxford Research Encyclopedia of Climate Science*. Oxford University Press. doi: <http://dx.doi.org/10.1093/acrefore/9780190228620.013.730>
Multiple data sets on ZENODO and via DoE HPSS
- Arora et al. Biorxiv 2018, 2019 (we have uploaded this before)
- doi:10.1080/03050629.2019.1632304
- doi:10.21105/joss.01322
- doi:10.1017/psrm.2018.11
- doi:10.1109/ISI.2018.8587373
- Brogun Dmitry Y. Faculty Day presentation on Innovations in Teaching and Learning: Biological Surveillance and Discovery in the Metagenomics Discovery Challenge! May 16. CUNY, Brooklyn College, NY (Conferences, Seminars and Symposiums: Faculty Day Talk) 2019
Brogun Dmitry Y. Data Science Innovation At the Intersection of Biomedical Research and the Library, June 13-14. National Library of Medicine Conference, Bethesda MD (Conferences & Invited Talks: NIH/NLM/NCBI/MDC Awards Metagenomics Discovery Challenge Conference Presentation), 2018
- <https://mdc2undergrad.commons.gc.cuny.edu/>
- Brogun, Dmitry Y, Levi, Kyle M, Robert A. Edwards, Klenk, Juergen, Watson, Rayneisha, Long, Alyssa, Busby, Ben. "Biological Surveillance and Discovery by Undergraduate Students in the Microbial Metagenomics Discovery Challenge!" , 2019, *Genome Research*, under submission
- M.J. Hannon, Jr., A.F. Queiruga, S. Finsterle. Enhanced flux modeling using a moving least squares interpolation basis in iTOUGH2. Presented at the 2018 TOUGH Symposium, October 8-10, 2018. <https://drive.google.com/file/d/1CYrFOCutcB6Rilv2yESNXN5r4O5ejQ8e/view?usp=sharing>

- M.J. Hannon, Jr., M. Stuckman, C. Lopano, J. A. Hakala. Modeling trace metal release from Marcellus Shale drill cuttings exposed to atmospheric conditions. Presented at the 2018 TOUGH Symposium, October 8-10, 2018.
https://drive.google.com/file/d/1pGBB4AduKkzOAqV9BCyKnBNyJzAysK_7/view?usp=sharing
- Arora, Rohit, Harry M. Burke, and Ramy Arnaout. "Immunological Diversity with Similarity." bioRxiv (2018): 483131.
- Arora, Rohit, et al. "Repertoire-Based Diagnostics Using Statistical Biophysics." bioRxiv (2019): 519108
- D. Rorabaugh, M. Guevara, R. Llamas, J. Kitson, R. Vargas, and M. Taufer. SOMOSPIE: A modular SOil MOisture SPatial Inference Engine based on data-driven decisions. 2019 15th International Conference on eScience, presented 25 September 2019.
- Software automation for research and training in neural engineering, <https://www.abstractsonline.com/pp8/#!/7883/presentation/71984>
- ACM does a much better job summarizing this than I:
 - <https://dl.acm.org/citation.cfm?id=3355752>
 - <https://dl.acm.org/citation.cfm?id=3333260>
 - <https://dl.acm.org/citation.cfm?id=3333151>
 - <https://dl.acm.org/citation.cfm?id=3101226>
 - <https://dl.acm.org/citation.cfm?id=3333253>
- <http://js-170-95.jetstream-cloud.org:3838/>
- The following lecture series (starting at UNIX) directs users to a terminal with an in-house webshell based on Shellinabox. It is provided as a free online web textbook. The webshell/terminal application runs on Jetstream.
 - Table of Contents URL: <http://charmm-gui.org/lecture>
 - First Lesson URL: <http://charmm-gui.org/lecture/unix>
 - Title: CHARMM-GUI Lecture Series
 - Authors: Nathan Kern, Wonpil Im, Jeffery Klauda, Patrick Fleming
 - Publication: Self-published on charmm-gui.org
- Hardy T., Atkins C., and Mentewab A., 2019. Diversity and prevalence of antibiotic resistance genes in plant root microbiomes. Spelman Research Day, April 25th, 2019, Spelman College, Atlanta, GA.
- Wright, D. W, Jeong, C.,Brookes, E., Zhang, H., Perkins, S. J., and Curtis, J. E., PDB Scan and PDB Rx: Easy to use tools to prepare initial models for molecular simulation and modeling. In preparation.
- Bowerman, S., Curtis, J. E., Clayton, J., Brookes, E. and Wereszczynski, J., (2019) BEES: Bayesian Ensemble Estimator from SAS. Biophys. J. 117:3 p 399-407. DOI:10.1016/j.bpj.2019.06.024
- Brookes, E., Stubbs, J. (2019). GenApp, Containers and Abaco. PEARC19, ACM, New York. Article 12, 8 pages. DOI: 10.1145/3332186.3332191
- Savelyev, A, Brookes, E. Plotting Advancements to the GenApp Framework. Gateways 2018. Austin, Texas. Presented.
- Brookes, E. (2018). Definition File Generation of Science Gateways using GenApp. International Workshop on Science Gateways. Edinburgh, Scotland, UK.
- Pierce, M., Miller, M., Brookes, E., Wong, M., Liu, Y., Afgan, E., Gesing, S., Dahan, M., Marru, S. and Walker, T. (2018). Towards a Science Gateway Reference Architecture. International Workshop on Science Gateways. Edinburgh, Scotland, UK.

- E. H. Brookes, J. P. Vachette, C. Jeong, M. Rocco. Advances to the UltraScan SOLUTION MOdeler (US-SOMO) SAS and HPLC-SAXS modules. XVII International Small Angle Scattering Conference. October 7-12 2018, Traverse City Michigan.
- Brookes, E., Stubbs, J. (2019). GenApp, Containers and Abaco. PEARC19, Chicago, July 2019.
- Chen, Y., Jeong, C., Savelyev, A., Krueger, S., Curtis, J. E., Brookes, E.H., Fushman, D. (2019) ROTDIF-web and ALTENS: GenApp-based Science Gateways for Biomolecular Nuclear Magnetic Resonance (NMR) Data Analysis and Structure Modeling, Gateways 2019, San Diego. DOI: 10.17605/OSF.IO/T4GKH
- Chourasia, A., Nedau, D., Luo, J., Chen, T., Miller, M., Brookes, E. (2019) Enabling rich data sharing for Science Gateways via the SeedMeLab platform. Gateways 2019, San Diego. DOI: 10.17605/OSF.IO/WV3BF
- Wilkins-Diehr, N., Miller, M., Brookes, E.H., Arora, R., Chourasia, A., Calyam, P., Jennewein, D.M., Nandigam, V., LaMar, M.D., Cleveland, S.B., Newman, G., Wang, S., Zaslavsky, I., Cianfrocco, M.A., Ellett, K., Tarboton, D. (2019) Measuring Success: How Science Gateways Define Impact. Gateways 2019, San Diego.
- Brookes, E., Savelyev, A. (2018). Creating Science Gateways using Advanced GenApp Capabilities / Hands-on workshop. Gateways 2018. Austin, TX.
- Brookes, E., Perkins, S. (2019). SASSIE2. 24th International Analytical Ultracentrifugation Workshop. Christchurch, NZ.
- Brookes, E., Jeong, C., Stubbs, J. (2019). Creating Science Gateways using GenApp, Containers and Abaco / Hands-on workshop. Gateways 2019. San Diego, CA.
- Leung, N. Brookes, E. Curtis, J. E. (2018). Poster: JSSSE: An Extensible and Configurable Molecular Structure Editor for Scientific Gateways. Gateways 2018. Austin, TX
- Chen, Y. Savelyev, A. Brookes, E. Fushman, D. (2018). Poster: ROTDIF-Web: A GenApp Generated Science Gateway for Comprehensive Analysis of Biomolecular NMR Relaxation Data. Gateways 2018. Austin, TX
- Brookes, E., Curtis, J.E., Fushman, D., Krueger, S., Savelyev, A. (2018) The GenApp Framework for SAS and Beyond. XVII International Small Angle Scattering Conference. Traverse City, MI.
- Chourasia, A., Luo, J., Nedau, D., Chen, T., Miller, M., Brookes, E. (2019) Enabling rich data sharing for Science Gateways via the SeedMe2 platform. Poster. PEARC19. Chicago, IL.
- Christie, M., Owczarzak, M., Ho, Y-S., Parsons, P., Krueger, S., Curtis, J.E., Brookes, E.H. (2019) GenApp Style and Layout Enhancements. Poster. Gateways 2019. San Diego, CA. DOI: 10.17605/OSF.IO/NQSPR
- Jeong, C., Vachette, P., Brookes, E.H. (2019) MultiSAXSHub: A GenApp-Generated Science Gateway for Unified Computational SAXS Modeling Tool. Poster. Gateways 2019. San Diego, CA. DOI 10.17605/OSF.IO/354CJ
- Guan, A., Casertano, M., Jeong, C., Brookes, E.H., Fushman, D. (2019) PARNMR: A GenApp-Generated Science Gateway for Comprehensive Analysis of Paramagnetic Effects in Biomolecular NMR Data. Poster. Gateways 2019. San Diego, CA. DOI: 10.17605/OSF.IO/92VCY
- <https://pb.appcivist.org/>
- Knox, E. B., and C. Li. 2017. The East Asian origin of the giant lobelias. *American Journal of Botany* 104: 924-938.

- Chafin TK, Douglas MR, Martin BT, Douglas ME. 2019. Hybridization drives genetic erosion in sympatric desert fishes of western North America. *Heredity* (20 August 2019): DOI 10.1038/s41437-019-0259-2
- bioRxiv 692509; doi: <https://doi.org/10.1101/692509>
- Chafin TK, Papudeshi B, Sanders S, Ganote C, Reshetnikov AN, Sokolov SG, Doak TG, Pummill JF, Douglas MR, Douglas ME. Identifying genomic underpinnings of diversification rate shifts in tapeworms using phylogenetic rate tests. Plant and Animal Genome Conference (PAG XXVII), San Diego CA, 12-16 January 2019.
- Papudeshi B, Chafin T.K., Sanders S, Ganote C, Reshetnikov AN, Sokolov SG, Doak TG, Pummill JF, Douglas MR, Douglas ME. De novo assembly of a novel tapeworm genome (*Nippotaenia percotti*) to examine genetic adaptations to parasitism in fish. Plant and Animal Genome Conference (PAG XXVII), San Diego CA, 12-16 January 2019.
- Martin BT, Douglas MR, Chafin TK, Placyk JS, Birkhead RD, Phillips CA, Douglas ME. 2019. Introgression and selection in *Terrapene* from the Southeastern United States. Box Turtle Conservation Workshop, Haw River State Park, NC, 17-19 May 2019.
- Martin BT, Douglas MR, Chafin TK, Placyk JS, Birkhead RD, Phillips CA, Douglas ME. 2019. Introgression and selection in *Terrapene* from the Southeastern United States. Joint Meeting of Ichthyologists and Herpetologists, Snowbird, UT, 22-28 July 2019
- Mussmann SM, Douglas MR, Douglas ME. 2019. Delimitation of genetic diversity in *Rhinichthys osculus* of the Owens and Amargosa River basins: Implications for biodiversity conservation and management. American Fisheries Society / The Wildlife Society Joint Meeting, Reno, NV, 29-Sept - 3-Oct. 2019
- Martin BT, Douglas MR, Chafin TK, Placyk JS, Birkhead RD, Phillips C, Douglas ME. 2019. Selection Along Ecological Gradients: Genome-Wide Assessment of Introgression and Adaptive mRNA Loci. American Fisheries Society / The Wildlife Society Joint Meeting, Reno, NV, 29-Sept - 3-Oct. 2019
- Zbinden ZD, Douglas MR, Douglas ME. 2019. Community Genomics of Stream Fish: A Framework for Multi-Species, Comparative Population Genetics Spanning the White River Basin. American Fisheries Society / The Wildlife Society Joint Meeting, Reno, NV, 29-Sept - 3-Oct. 2019
- Douglas MR, Chafin TK, Martin BT, Zbinden ZD, Gray C, Middaugh C, Ballard J, White D, Douglas ME. 2019. Population Connectivity and Chronic Wasting Disease Susceptibility of White-Tailed Deer in Arkansas: A Landscape Genomics Approach. American Fisheries Society / The Wildlife Society Joint Meeting, Reno, NV, 29-Sept - 3-Oct. 2019
- Chafin TK, Douglas MR, Douglas ME. Is the Red Wolf a Hybrid Species? Revisiting Genomic Data Using Coalescent Predictions. American Fisheries Society / The Wildlife Society Joint Meeting, Reno, NV, 29-Sept - 3-Oct. 2019.
- Wangmo S, Wangchuk K, Douglas MR, Claussen JE, Philipp DP, Douglas ME. 2019. So Much Noise! Extracting Tag Signals from Radio-Telemetry Data Using a Custom Python Script. American Fisheries Society / The Wildlife Society Joint Meeting, Reno, NV, 29-Sept - 3-Oct. 2019. <poster>
- Papudeshi B, Chafin TK, Sanders SA, Ganote C, Reshetnikov AN, Sokolov SG, Doak TG, Pummill JF, Douglas MR, Douglas ME. 2019. Genome and transcriptome analysis of fish tapeworm: *Nippotaenia mogurndae* through scientific collaboration between international research labs and U.S. cyberinfrastructure. American Fisheries Society / The Wildlife Society Joint Meeting, Reno, NV, 29-Sept. 3-Oct. 2019 <poster>

- Tchoua, Roselyne B., Aswathy Ajith, Zhi Hong, Logan T. Ward, Kyle Chard, Alexander Belikov, Debra J. Audus, Shrayesh Patel, Juan J. de Pablo, and Ian T. Foster. "Creating Training Data for Scientific Named Entity Recognition with Minimal Human Effort." In International Conference on Computational Science, pp. 398-411. Springer, Cham, 2019.
- Haber, Jaren. 2020. Sorting Schools: A Computational Analysis of Charter School Identities and Stratification. Forthcoming in "Sociology of Education".
- Haber, Jaren. 2019. School Websites and Racial Meanings: Evidence from Computational Text Analysis and Online Experiments. Round-table presentation at American Sociological Association Annual Meeting.
- Shepherd, T.J., Barthelmie, R.J., and Pryor, S.C. (2019): Sensitivity of wind turbine array downstream effects to the parameterization used in WRF. Journal of Applied Meteorology and Climatology.
- Pryor, S.C., Shepherd, T.J., Volker, P., Hahmann, A.N., and Barthelmie, R.J. (2019): Wind theft from onshore arrays: Sensitivity to wind farm parameterization and resolution. Journal of Applied Meteorology and Climatology.
- Shepherd, T.J., Barthelmie, R.J., and Pryor, S.C. (2019): Assessment of Wind Turbine Impact on Future Climate in GCM-Driven WRF Simulations. Journal of Physics: Conference Series.
- Letson, F., Shepherd, T.J., Barthelmie, R.J., and Pryor, S.C. (2019): Modelling Hail and Convective storms with WRF for Wind Energy Applications. Journal of Physics: Conference Series.
- Pryor, S.C., Shepherd, T.J., Bukovsky, M., and Barthelmie, R.J. (2019): Assessing the stability of wind resource and operating conditions. Journal of Physics: Conference Series.
- Lee, M.D. (2019). GToTree: a user-friendly workflow for phylogenomics. Bioinformatics. <https://doi.org/10.1093/bioinformatics/btz188>
- Lee, M.D. et al. (2019). Marine Synechococcus isolates representing globally abundant genomic lineages demonstrate a unique evolutionary path of genome reduction without a decrease in GC content. Environmental Microbiology. <https://doi.org/10.1111/1462-2920.14552>
- I submitted my work, but no response for now
- <https://www.ncbi.nlm.nih.gov/pubmed/31097508>
- 10.1007/978-1-4939-8775-7_9
- <https://www.nature.com/articles/s41597-019-0073-y>
- <https://doi.org/10.25663/brainlife.pub.3>
- <https://doi.org/10.25663/brainlife.pub.4>
- <https://doi.org/10.25663/brainlife.pub.5>
- <https://doi.org/10.25663/brainlife.pub.8>
- Gourav Rattihalli, Madhusudhan Govindaraju, Devesh Tiwari: Towards Enabling Dynamic Resource Estimation and Correction for Improving Utilization in an Apache Mesos Cloud Environment. CCGRID 2019: 188-197
- N/A