Jetstream2 NSF Annual Report
October 1, 2022 – September 30, 2023

1. Accomplishments

1.1. What are the major goals of the project?

The purpose of the Jetstream2 computational resource is to ensure that the science and engineering community has ready access to the advanced computational and data-driven capabilities required to tackle today's most complex research problems and issues. Jetstream2 is a computational research resource for the national open (unclassified) research and education (R&E) community - a data analysis and computational resource that US scientists and engineers will use interactively as well as programmatically through advanced interfaces. This system has already begun to help researchers and educators make new discoveries and train the next generation workforce. The project team will continue to innovate, accelerate, and provide on-demand services for the R&E community throughout the US who will find Jetstream2 to be novel, fast, and reliable. Jetstream2 now provides innovative and widely needed services that complement and add to the existing, and future, portfolio of resources within the NSF Innovative High Performance Computing (HPC) Program.

The primary project deliverable of the Jetstream2 award is the Jetstream2 system. The goal of this report is to describe the details of acceptance processes to ensure the system is a valuable resource for researchers and educators, and meets the agreed upon terms and vision proposed in conjunction with the NSF and partners. Jetstream2 is a fully-configurable large-scale cloud computing resource that leverages both on-demand and persistent virtual machine technology to support a wide array of software environments and services. As the open cloud for science, Jetstream2 provides:

- A core cloud computing platform at Indiana University (IU) featuring 506 nodes (a mix of central processing unit (CPU), graphics processing unit (GPU), and larger-memory) with a processing capability of 9 peta floating point operations (PFLOPS) attached to 14.2 PB of hybrid-storage.
- Four regional cloud systems with an aggregate processing capacity of 632 tera floating point operations (TFLOPS) from 32 CPU nodes, 6 GPU nodes, and 3.1 petabyte (PB) storage, hosted at Arizona State University (ASU), Cornell University, the University of Hawai'i (UH), and the Texas Advanced Computing Center (TACC).

- A hybrid high-performance storage environment with high input/output operations per second (IOPS) flash media and high-capacity disk drives to support data-intensive and data-distribution endeavors.
- A user-centric, researcher-friendly, self-service interface and OpenStack Infrastructure-as-a-Service software to enable interactive data analyses.
- Support for programmable cyberinfrastructure and interoperability with commercial clouds with these features:
  - Native container support
  - Integration of orchestration tools with support for cross-platform deployment
  - Graphical interactive environments
  - Push-button virtual clusters and elastic capability
  - Highly available and persistent science gateway hosting services
  - Federated authentication for JupyterHubs

1.2. What was accomplished under these goals?

1.2.1. Major Activities

The major activities for this project are described below:

- The system met or exceeded uptime and completion requirements per the goals in our project execution plan, ≥99% uptime and ≥96% job completion (see Section 1.2.3).
- The project completed the first annual community user survey among 2,168 individuals. All key findings can be found in Section 1.2.4.
  - Overall, 96.4 percent of all respondents report being “satisfied” or “extremely satisfied” with Jetstream2’s performance, up from the 89.3% reported for Jetstream1 in 2021.
  - In 2023, all service areas of the evaluated scored above 4.0 (on a 5.0 scale) in terms of user satisfaction.
  - Individuals report the highest levels of satisfaction with quality of response to questions via help@jetstream-cloud.org (4.64/5.0) and with speed of response to questions via help@jetstream-cloud.org (4.60/5.0). Users also reported a high mean satisfaction with the speed (responsiveness) of Jetstream2 (4.56/5.0), the speed of response to questions via the ACCESS ticket portal (4.56/5.0), and the quality of responses to questions via the ACCESS ticket portal (4.56/5.0).
  - Just over ninety-two (92.2) percent of respondents, up from the 89.8 percent for Jetstream1 in 2021, indicate that Jetstream2 is “very important” or “essential” to their research activities, with a mean importance rating of 4.42 (on a 5.0 scale).
  - Nearly eighty-six (85.7) percent of respondents, up from the 83 percent reported for Jetstream1 in 2021, indicate the resource is “very important” or “essential” to their education activities, with a mean importance rating of 4.22 (on a 5.0 scale).

- Major operating system updates for the host system were performed during this project period (a feature rarely performed in traditional HPC environments) moving from Ubuntu 20.04 LTS to 22.04 LTS
- The OpenStack and Ceph environments were also updated on the core resource this year with experiences shared with regional systems.
- Increased monitoring for operations was added for many issues discovered as detailed in Section 1.2.2.
- Exosphere, the default interface for Jetstream2, has also undergone numerous enhancements related to storage volumes, sharing, and floating IP management as detailed below.
- During this reporting period the Jetstream2 environment has provided resources to over 370 projects, 1,800 unique individuals, and 65 science gateways (indirectly serving over 165K individuals). Details are provided in Section 1.2.3 below.
- The project team presented a virtual review covering the first year of operations on April 25, 2023.
- The University of California San Diego, specifically the San Diego Supercomputing Center, joined the Jetstream2 project during this project period to provide additional advanced computational support expertise and explore resource partnerships with CloudBank.
- The Jetstream2 team participated in the development of 16 journal articles and juried conference productions in addition to software products and other presentation material produced. The extensive listing of all products can be found in Section 2.

Note that most data and findings in this report are included through August 2023 in order to submit material to the NSF before being flagged as overdue on September 30, 2023.

1.2.2. Specific Objectives:

Accomplishments relevant to the achievement of goals for this project are described below:

Indiana University

- Jetstream2 nodes were upgraded from Ubuntu 20.04 LTS to 22.04 LTS operating system and OpenStack services. Jetstream2 utilizes Ubuntu's autopatch facility whereby security patches are automatically applied to the hosts. Additionally, in Q3 2023, the Linux 6.2 kernel was deployed on the IU cloud, marking a major version upgrade from the default 5.x kernels. This upgrade should yield significant performance enhancements for AMD processors like the EPYC3 Milan CPUs in Jetstream2.
- Jetstream2 was upgraded to OpenStack Zed in mid 2023. This is the current OpenStack release running on Jetstream2. Ceph was also upgraded to the Quincy release during this reporting period.

- The image build pipeline integrated Rocky 9 and Alma 9 and removed support for Ubuntu 18 and CentOS 7. Images are updated weekly, with security updates loaded prior to a new instance launch unless specifically disabled at runtime.
- Prometheus monitoring has been deployed for services on the Jetstream2 cloud, including watching for downed hosts, Ceph issues, abnormally high utilization on the control plane, and other monitoring. This continues to be refined and built on as needs arise. Efforts have started to put these monitoring services on an external, dedicated host on the Indiana University network.
- Exosphere, the default user interface for Jetstream2, has undergone many usability and feature improvements critical to system acceptance and operations, those include:
    - Preliminary support for Manila shares (Filesystems-as-a-service)
    - Support for retaining a floating ip for future use upon instance deletion
    - Support for instance operating systems: Rocky Linux and AlmaLinux v9
    - Showing volume snapshots in volume quota usage
    - Adding additional confirmation for bulk instance deletion
    - Showing status of snapshot creation as it progresses
    - Added the ability to share images with other projects and the community
    - Many more user interface refinements

TACC

- OpenStack software upgraded and managed to match IU (OpenStack Zed).
- TACC nodes were upgraded from Ubuntu 20.04 LTS to 22.04 LTS operating system.
- Contributed to Jetstream2 code repository with configuration information critical to regional system deployments.
- Jetstream2 was the primary platform for the University of Texas at Austin's Computational Engineering Software Systems Design (COE 332) Spring 2023

University of Arizona

CACAO included several enhancements:

- support for user-defined, non-openstack cloud providers that can be used privately or shared with others (backend-only)
- metadata: redesigned cacao template metadata, including metadata for dynamic UI wizard and documentation links
- credentials: support for private ssh keys
- image selection for deployments now honors min_disk and min_ram when displaying flavors
- rebuilt guacamole ui component to better support file transfers
- added a maintenance banner
- template changes:
    - added boot volumes for some templates
    - support for templates in private git repos
    - Terraform templates now use http-backend state storage
- New templates templates added:

    - NIH DADI
    - multi-node docker swarm
    - vms4workshop, template to launch multiple configured vms with large workshops
- Jupyterhub template: Added shared storage
- CACAO UI-specific changes
    - added a template catalog page
    - added deployment editing
- Authentication: added support for linked CILogon/ACCESS identities

Arizona State University

- Ubuntu OS has been updated from 20 to 22.
- OpenStack software has been upgraded to match IU (OpenStack Zed).
- VM instances have been tested through Exosphere.
- Monthly broad Research Computing training events give an overview of JS2. There have been 23 events with 266 attendees since the last reporting period.
- A JS2-focused workshop is planned for early fall 2023
- Moving ASU JS2 equipment to new data center late fall 2023

Johns Hopkins University (JHU)

- Continued to leverage and maintain support for submitting user jobs from usegalaxy.org science gateway to Jetstream2. Over the period, 662,681 jobs were submitted on behalf of 26,657 users.
- Added support for running GPU jobs from usegalaxy.org server, and 225 GPU jobs were run to date.
- Ran benchmarks for RNA-Seq and variant calling workflows as two of the most popular types of workloads executed on usegalaxy.org. The benchmark results will be used to better estimate resource requirements for tools and allow users to better understand resource requirements when they run such workloads on commercial cloud providers.

University of Hawai'i

- Added ASC190026 - Hawaii EPSCoR - Change-HI Project (NSF#OIA-2149133) to the Hawaii region. This project now leverages both IU and UH regions to support the EPSCoR projects data ingestion and aggregation pipelines for the Hawaii Climate Data Portal (https://hawaii.edu/hcdp).
- Added CIS230142 - National Vulnerability Intelligence Platform (NVIP),a Department of Homeland Security project related to cybersecurity for software vulnerability assessment -deployed to the UH region .
- Added CIS230179 - Tapis Framework project for UH to region (NSF #1931575, #1931439) to support UH science gateway projects.
- Added CIS230039 - Drone Photogrammetry to 3d mesh & point-cloud products for terrestrial data project to Help Dr. Romina King from the University of Guam.

- Added CIS230038- Developing a proof-of-concept Service/Gateway to look at 3d images/meshes/point-cloud data for outreach and annotation for UH Lava lab collaboration with University of Guam.
- Used JS2 for during Smart Sensor Data Workshop in March 2023 - 13 attendees for Cyberinfrastructure Training to Advance Environmental Science(NSF #2118222)
- Smart Amplified Group Environment (SAGE3) deployed successfully and running in JS2 (NSF #2004014, #2003800, #2003387)
- Gave workshop on Jetstream2 to Hawaii EPSCoR project group ~40 attendees.
- Leveraged JS2 to support Change-HI summer REU student CI research projects.
- Regular updates were performed to the Ubuntu 20.04 LTS operating systems and OpenStack services.

University Corporation for Atmospheric Research (UCAR)

- The Unidata program at UCAR successfully migrated the Unidata Science Gateway, including all of the data and software services to JS2.
- In addition, Docker and Kubernetes were also transitioned to JS2 in order to support scaling for a larger number of users.
- Unidata is also experimenting with the use of JS2 computing infrastructure, including both CPU and GPU instances, to deploy the Weather Research and Forecast (WRF) model. Unidata staff are exploring Jetstream2 GPU hardware to deploy advanced deep learning in Earth Systems Science. In collaboration with Andrea Zonca, SDSC, we have deployed a GPU-enabled JupyterHub currently hosted at https://jupyterhub.unidata.ucar.edu equipped with TensorFlow and PyTorch software as well as geoscience notebooks that employ this specialized hardware.
- Both Jetstream and Jetstream2 resources have been of tremendous help in supporting the Unidata community, including supporting many training workshops at universities by providing customized JupyterHub servers at an accelerating pace. Since the launch of Jet- stream2, we have deployed PyAOS (Python for Atmosphere and Ocean Science) multi-user JupyterHub installations tailored to the requirements of instructors at 12 universities. To date, 449 students have used Unidata JupyterHub servers running on Jetstream2 in this manner1. In addition, Jetstream2 resources were crucial to Unidata's activities during the American Meteorological Society (AMS) 2023 annual meeting. Unidata supported three workshops with customized PyAOS JupyterHub servers: a MetPy Short Course, Student Python Workshop, and the Colorado State University LROSE radar meteorology workshop. In all, Unidata had 141 participants using Jetstream2 resources at AMS 2023.
- Held from June 5-8, the 2023 Unidata Users Workshop3 invited attendees to explore effective, ethical, and reproducible methods for communicating scientific insights using Earth Systems Science data. The event, which covered four main themes — storytelling with data; multidisciplinary data usage; ethical application of AI/ML; and open, reproducible science — emphasized practical outcomes, including data navigation, visualization, scientific storytelling, AI/ML applications, and contributions to open science. Many sessions employed live Jupyter notebooks, enabling participants to

follow along interactively, engage in take-home exercises, and extend their exploration beyond the workshop's conclusion. Thanks to Jetstream2 resources, Unidata was able to enhance the workshop experience by providing PyAOS and AI/ML-ready JupyterHub environments to 66 attendees.

- Unidata continues to host publicly accessible AWIPS EDEX servers on the Jetstream2 cloud platform, where we serve real-time meteorological and geographic data to AWIPS CAVE clients at several dozen universities. Our EDEX servers are also accessible via the python-awips data access framework API. We continue using Jetstream2 to develop cloud-deployable AWIPS instances as virtual machines images available to users of the OpenStack CLI.
- Another application of Jetstream2 resources was its use to provide simulation capabilities to users. In collaboration with Greg Blumberg at Millersville University, Unidata staff have deployed a single- column WRF model in a JupyterHub environment for undergraduate instructional objectives

Cornell University

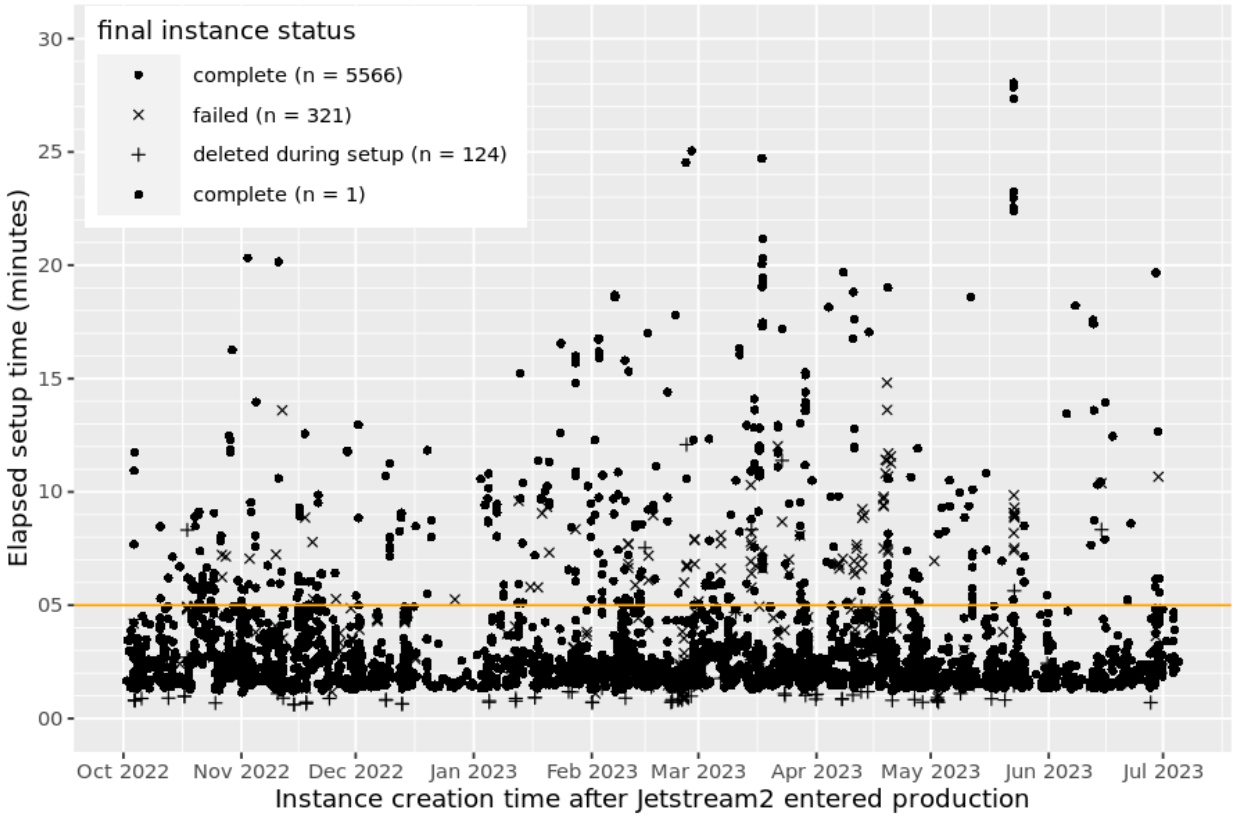
- Worked with Cornell Epigenomics Core (Pugh Lab) in order to provide workflow execution on Cornell zone of JS2
- Engaged Bronx HS of Science student in mentorship program to look at OpenXDMoD monitoring for virtual clusters on Cornell zone
- Prepared Sara Pryor (Cornell Earth and Atmospheric Sciences) for transition to Cornell zone
- Applying lessons learned during Cornell zone set up to implementation on new version of local OpenStack cloud, Red Cloud
- Added BIO220060 to Cornel Region - ly349@cornell.edu - https://yulab.org
- Resa Reynolds added to Jira to assist with tickets where appropriate

1.2.3. Significant results

We have the following significant technical results so far:

Uptime and launch success are two key performance indicators for Jetstream2. The detailed outage information is shown in Section 5.2 in the outages tables. Below are summaries of the results.

- Since entering operations, there have been 62.35 hours of outages total where cloud services were partially or fully impaired.
- Out of 6,528 hours of operations since 10-1-2022, this is 0.447% downtime for production operations to date. This yields an uptime of 99.04% with the performance expectation from the solicitation being 95% or greater and our own increased goal of 99% or greater uptime.
- Launch statistics for featured images were measured in early operations and continue to be measured into operations
    - Launches were successful 96.02% for production operations (OY1 Q1-Q3 / PY3 Q1-Q3 2022-10-01 to 2023-07-05). This exceeds the goal set in the PEP.



Additional statistics of note:

	Established goals from PEP	Pre-Ops March 2022	PY2 Q3 Apr - Jun 2022	PY2 Q4 July - Sept 2022	PY3 Q1 Oct - Dec 2022	PY3 Q2 Jan - Mar 2023	PY3 Q3 April - June 2023	July 2023	August 2023
Capacity of system allocated via NSF-specified allocation process	90%	>97%	>97%	>97%	>97%	>97%	>97%	>97%	>97%
Total number of distinct users	1000 Users (OY 1)	348	1,144	1,700	1,844	1,882	1,850	1,713	1,830
Total number of students having used Jetstream in an educational or training setting	350 (OY 1)	127	391	568	611	499	375	336	309

Total number of science gateways using Jetstream (active total)	45 (OY 1)	14	15	51	53	56	57	61	65
Total number of users served via Science Gateways**	75,000 (OY 1)	68,122	69,556	169,210	169,210	169,210	169,210	169,210	169,210
SUs available to user community per month	Baseline *	25M CPU 5M GPU 3.5M LM	25M CPU 5M GPU 3.5M LM	25M CPU 5M GPU 3.5M LM	25M CPU 5M GPU 3.5M LM	25M CPU 5M GPU 3.5M LM	25M CPU 5M GPU 3.5M LM	25M CPU 5M GPU 3.5M LM	25M CPU 5M GPU 3.5M LM
Active Projects	Maintain 275+	113	245	322	357	333	373	362	368
Intitutions Represented	Maintain 250+	102	241	283	269	263	259	250	246
Fields of Science	Maintain 55+	31	60	66	65	67	71	73	74
* Baseline that allows maximum available SUs while maintaining resources to be "on demand"									
** Updated when allocations renew, generally so may be fairly static									

Allocation statistics for September 2022 to August 31, 2023:

Jetstream2 continues forward with ACCESS allocations, transitioning XSEDE-era allocations to ACCESS-era allocations as they come up for renewal. PY3 allocations continue the trend of more modest allocation sizes when compared to national cyberinfrastructure HPC system with ACCESS Explore (up to 400,000 credits) and Discover (up to 1.5M credits) consisting of the majority of allocations. This is consistent with the mission of Jetstream2 to serve research projects and education that are not well served by traditional HPC.

	XSEDE Totals (Jan 2022 - Aug 2022)		**9/1/2022 - Pre-Ops Q3/Ops Q3 ACCESS (PY2 Q4.3) through August 2023 (OY1 Q4.2 / PY3 Q4.2)**				**ACCESS Total to Date Sep 2022 to Present**	
Resource	**SUs**	**Storage**	**JS2 Compute**	**JS2 GPU**	**JS2 LM**	**JS2 Storage**	**SUs**	**Storage**
	Startup		***Explore***				***Explore***	
Total requests	108	70	131	49	8	62	188	62

SUs awarded	38,350,000	282,874	15,138,540	5,147,798	926,000	571,300	21,212,338	571,300
	Education		***Discover***				***Discover***	
Total requests	38	24	62	33	14	40	109	40
SUs awarded	21,982,000	107,000	21,823,195	8,577,000	3,320,000	362,950	33,720,195	362,950
	Campus Champion/Staff		***Accelerate***				***Accelerate***	
Total requests	111	36	8	4	0	6	12	6
SUs awarded	50,859,000	34,500	5,401,090	959,664	0	178,000	6,360,754	178,000
	Supp/Disc/Net Xfer		***Maximize***				***Maximize***	
Total requests	92	71	11	4	1	7	16	7
SUs awarded	50,740,673	320,800	4,027,904	2,500,000	2,250,000	64,000	8,777,904	64,000
	Research		***Supplemental/Discretionary/Net Transfers/Exchanges***				***Supplemental/Discretionary/Net Transfers/Exchanges***	
Total requests	62	30	94	59	20	48	173	48
SUs awarded	43,453,918	510,506	21,418,354	12,746,003	1,916,760	638,900	36,081,117	638,900
	Total requests and allocations		***Total requests and allocations***				***Total requests and allocations***	
Total requests	**411**	**231**	**306**	**149**	**43**	**163**	**498**	**163**
Total SUs awarded	**205,385,591**	**1,255,680**	**67,809,083**	**29,930,465**	**8,412,760**	**1,815,150**	**106,152,308**	**1,815,150**

This reporting period completed the transition from XSEDE allocation types to ACCESS allocation types. We have not included historical data as it would require having a separate set of tables denoting XSEDE allocations since they do not precisely map to the new allocations style.

1.2.4. Key outcomes or Other achievements

The first annual Jetstream2 user survey[1] was completed during this period with the following key findings:

- Overall, 96.4 percent of all respondents report being “satisfied” or “extremely satisfied” with Jetstream2’s performance, up from the 89.3% reported for Jetstream1 in 2021. Applying a standard Likert scale to the responses offered, with “1” being “extremely dissatisfied” and “5” being “extremely satisfied,” the mean satisfaction is 4.60 on a 5.0 scale.
- In 2023, all service areas of the evaluated scored above 4.0 (on a 5.0 scale) in terms of user satisfaction.
- Users report the highest levels of satisfaction with quality of response to questions via help@jetstream-cloud.org (4.64/5.0) and with speed of response to questions via help@jetstream-cloud.org (4.60/5.0). Users also reported a high mean satisfaction with the speed (responsiveness) of Jetstream2 (4.56/5.0), the speed of response to questions via the ACCESS ticket portal (4.56/5.0), and the quality of responses to questions via the ACCESS ticket portal (4.56/5.0).
- Documentation about Jetstream2 was the lowest-rated item in the 2023 assessment (4.12 /5.0).
    - Among dissatisfied users, the most prevalent comments involve deficiencies with documentation.
- Just over ninety-two (92.2) percent of respondents, up from the 89.8 percent for Jetstream1 in 2021, indicate that Jetstream2 is “very important” or “essential” to their research activities, with a mean importance rating of 4.42 (on a 5.0 scale).
- Nearly eighty-six (85.7) percent of respondents, up from the 83 percent reported for Jetstream1 in 2021, indicate the resource is “very important” or “essential” to their education activities, with a mean importance rating of 4.22 (on a 5.0 scale).
- The ability to access computational resources on-demand was cited as the most useful characteristic of Jetstream2 by 24 percent of respondents, followed closely by the Exosphere user interface at 22 percent.
- Users were largely neutral to positive about the training methods they were asked to rate and, as in previous years, respondents showed a clear preference for self-service through just-in-time, online resources.
- Users from minority-serving institutions and/or EPSCoR states comprise 13 percent of the Jetstream2 user population, up from 9 percent reporting in 2021 Jetstream1 assessment.

[1] https://scholarworks.iu.edu/dspace/handle/2022/29415

1.3. What opportunities for training and professional development has the project provided?

The Jetstream2 team has provided extended consultations to several universities and research organizations in order to help further research impacts as detailed in Section 1.4. This work has been completed through participation in the Midwest Research Computing & Data Consortium, the OpenInfra Summit, and the ACCESS Operations (CONECT project's) STEP program.

1.4. How have the results been disseminated to communities of interest?

Discussions and ongoing consultations for broader impacts

- Bi-weekly meeting with University of Alabama Birmingham to discuss cloud architecture and best practices. Discussions are ongoing for a potential Jetstream2 regional cloud at their site.
- Had initial discussions with University of Illinois Urbana Champaign about cloud architecture in March 2023 with a follow-up visit to IU in August 2023. Further discussions and in-person meetings are being planned.
- Gave a brief talk about Jetstream2 resources at the Midwest Research Computing & Data Consortium meeting in March 2023 to other research support staff in the region.
- Consulted with Georgia State University HPC/Cloud staff at the OpenInfra Summit and made plans for longer term consultations on cloud architecture and best practices.

Meetings with research groups interested in being future users of Jetstream2:

- Presented Jestream2 to members of the Matthew Hahn lab at IU as a potential computational resource for their analysis and processing pipelines.
- Discussed Jetstream2 with Sagar Samtani, Assistant Professor at Indiana University, as a potential solution for his Operations and Decisions Technologies lab's research. - October 2022
- Met with representatives of the FABRIC project (https://fabric-testbed.net/) to discuss collaboration with Jetstream2 - November 2022
- Spoke with members of Enterprise Computing at Indiana University in regards to Wiehe Lab at IU about their research needs and whether migrating to Jetstream2 or other resources from a campus VMWare deployment was a viable option - January 2023
- Discussed new research opportunities with Kevin Tyle and Brian Rose about an NSF/EarthCube grant (Project Pythia - https://www.earthcube.org/project-pythia) and how Jetstream2 might work for this project - February 2023
- Renewed discussions with Dr. Mark Frank for a previously stalled project on using AI for lung cancer detection and diagnosis. - April 2023

- Had conversations with Han Lab at Texas A&M University about science gateway hosting and other potential research needs - April 2023
- Continued discussions with Drs. Tang and Wang from Indiana University about Secure Virtualization environments for their NSF-funded research under the Center for Distributed Confidential Computing (CDCC) - April 2023
- Discussed Jetstream2 with Professor Sha Cao from Indiana University as a potential solution for the Biomedical Research Data Lab for their Alzheimer's disease and related dementia processing pipeline - April 2023
- Sean Cleveland of the Hawai'i regional site met with local faculty groups on using Jetstream2 for their research and education needs - April 2023

Courses and workshops utilizing Jetstream2 resources:

1. E631: Quantitative Research in Music Education - Indiana University
2. E632: Advanced Quantitative Research in Music Education - Indiana University
3. BIOL647 Digital Biology - Texas A&M University
4. BIOL350 Computational Genomics - Texas A&M University
5. BIOL650 Genomics - Texas A&M University
6. BIOL491/BIOT685 Directed Studies - Texas A&M University
7. BIS180L Genome Biology Lab - University of California Davis
8. BIO 351 Bioinformatics - Reed College
9. CompBio2023 - Advanced studies workshops in computational biology (NSF Award #1953405)
10. Population Genetics (Fall 2022/Spring 2023) San Diego State University
11. BIS 134 Systems Biology: From Biological Circuits to Biological Systems - University of California Davis
12. BIO 457 Fundamentals of Bioinformatics - SUNY Oswego
13. NASA JPL 2023 Preparing for Research in Space Microbiology (PRISM) Program - HBCU/MSI Workshop - California State University, Northridge; California State Polytechnic University, Pomona; Alabama State University; and Howard University
14. NASA GeneLab 2023 Amplicon Sequencing Workshop
15. BIO 404V/504V: Phylogenetics and Systematics - Arkansas State University
16. STAMPS 2023 Workshop - Marine Biological Laboratory Woods Hole
17. SDSC HPC Students: A program for educating and training the next generation of HPC professionals and researchers - San Diego State University
18. Chem Compute Undergraduate Computational Chemistry Science Gateway - multiple institutions - undergraduate instructional chemistry gateway - created by Sonoma State University
19. COS 442/542: Cloud Computing - University of Maine
20. B547/I533: Systems & Protocol Security & Information Assurance - Indiana University (Fall 2022/Spring 2023)
21. Deep Learning Tutorial for Translational AI Center - Iowa State University
22. IU-Purdue Student Cluster Competition Training Program - Indiana University and Purdue University

23. CS4663 Course of Distributed and Cloud Security - University of Texas San Antonio
24. ENGR 516 Engineering Cloud Computing - Indiana University
25. Cybersecurity Training (multiple workshops) - Center for Applied Cybersecurity Research - Indiana University
26. INFO 5502 / DTSC 5502 - Principles and Techniques for Data Science - University of North Texas
27. Containerization, application development, and large-scale data management - University of Texas San Antonio
28. COE 332 Software Engineering and Design - University of Texas
29. MSIS Capstone Program - Indiana University
30. ACCESS Student Training and Engagement Program - multiple organizations
31. Workshop in Molecular Evolution - Marine Biological Laboratory at Woods Hole
32. I520/B544 Security for Networked Systems - Indiana University
33. Research Workshop - University of Hawaii EPSCoR
34. Exploring Parallel Programming (graduate project) - California Institute of Technology
35. MATH 3399 Applications of Parallel Computer for Computational and Mathematical Biology - Jarvis Christian College
36. STAT 129 (Analyzing and Processing Big Data) - California State University-Sacramento
37. RMACC symposium/workshop for researchers (How to use national cyberinfrastructure) - Rocky Mountain Area Coordination Center
38. PHYS 477 Special Problems in Experimental and Theoretical Physics - Western Illinois University
39. PHYS 577 Special Problems in Physics - Western Illinois University
40. First year seminar class "AI for Everyone" - University of Mount Union
41. Genome Sequencing - North Central High School (Spokane, WA)
42. Marine Genomics - University of California Davis
43. BTEC395 "Translational Bioinformatics", and BTEC423 "Machine Learning with Bioinformatics Applications" - University of Maryland
44. Cell Evolution Simulator - educational resource for learners interested in the biological sciences - University of Arizona
45. WSSU Astrobotany lab bioinformatics workshop (HBCU) - Jet Propulsion Laboratory hosted workshop
46. Reinforcement Learning for Conservation Decisions - University of California, Berkeley
47. UAB Omics Hackathon 2023: Intelligent Safety - University of Alabama, Birmingham
48. CSCI 455: Principles of Database Systems - University of Puget Sound
49. Computational Social Science Methods (PA397C|INF385T) - University of Texas at Austin
50. Computational drug discovery course that will be taken by students from multiple institutions online - developed by Georgetown University
51. CMU LTI Courses 11-927 and 11-920 - Carnegie Mellon University
52. Computational Engineering Software Systems Design (COE 332) Spring 2023 - University of Texas at Austin

1.5. What do you plan to do during the next reporting period to accomplish the goals?

Technical highlights of this work will include:

- Create a new Jetstream2 Trial Allocation portal with the corresponding allocation infrastructure in a special trial allocations namespace
- Expand on the monitoring of all Jetstream2 systems to help improve response times and make a more robust and stable user experience
- Create both public and private facing analytics site to show researchers capacity and activity on the cloud and give administrators better insights into performance
- Increase the scope of the control plane to include an additional controller for added stability
- Continue regular updates of featured images and correct any automation issues.

Administrative aspects of this work will include:

- Monitoring and updating project risks.
- Auditing financial spending to ensure partners adhere to plans.
- Continuing collaborations with researchers and educators to provide letters for proposals to funding agencies.
- Engage in discussions where appropriate with institutions interested in deploying their own Jetstream2 regional cloud
- Further broader impacts by consulting with universities and colleges that run or want to run their own private cloud systems

User service and access elements of this work will include:

- Perform a document review and update out of date documentation
- Create onboarding documents for new students and researchers for allocations and the Exosphere interface
- Explore creating how-to videos for common tasks
- Continue to post and improve outage notification and status processes.

Education, outreach, and training aspects of this work will include:

- Sponsoring US-RSE (Research Software Engineering) conference, hosting an outreach table, and giving one of the keynote talks
- Participating in multiple CSCC events, Gateways23, SC23, and other targeted events and venus
- Developing asynchronous webinars for self-guided viewing
- Further discussions with various institutions as regards in-person events.

2. Products (resulting from this project during the specified reporting period)

Journal Articles and Juried Conference Papers

Jetstream project team generated (in bold). User-generated in plain text.

1. Olaya, P., Kennedy, D., Llamas, R., Valera, L., Vargas, R., Lofstead, J., & Taufer, M. (2022). Building trust in earth science findings through data traceability and results explainability. IEEE Transactions on Parallel and Distributed Systems, 34(2), 704-717.
2. Maghami, I., Van Beusekom, A., Hay, L., Li, Z., Bennett, A., Choi, Y., ... & Goodall, J. L. (2023). Building cyberinfrastructure for the reuse and reproducibility of complex hydrologic modeling studies. Environmental Modelling & Software, 164, 105689.
3. Suzumura, T., Sugiki, A., Takizawa, H., Imakura, A., Nakamura, H., Taura, K., ... & Uchibayashi, T. (2022, September). mdx: A cloud platform for supporting data science and cross-disciplinary research collaborations. In 2022 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech) (pp. 1-7). IEEE.
4. **Ranawaka, I., Abeysinghe, E., Wannipurage, D., De Silva, D., Brookes, E., Marru, S., ... & Pierce, M. (2023). Airavata Metascheduler: A Reliable, Fault Tolerant, and Resource-Aware Job Scheduling Service. In Practice and Experience in Advanced Research Computing (pp. 35-42).**
5. **Marru, S., Pierce, M., Plale, B., Pamidighantam, S., Wannipurage, D., Christie, M., ... & Yoshimoto, K. (2023). Cybershuttle: An End-to-End Cyberinfrastructure Continuum to Accelerate Discovery in Science and Engineering. In Practice and Experience in Advanced Research Computing (pp. 26-34).**
6. **Skidmore, E., Cosi, M., Swetnam, T., Merchant, N., Xu, Z., Choi, I., ... & Yung, M. (2023). Cloud Computing for Research and Education Gets a Sweet Upgrade with CACAO. In Practice and Experience in Advanced Research Computing (pp. 251-254).**
7. **Coulter, E., Knepper, R., Reynolds, R., Bird, S., Samuel, T. K., & Snapp-Childs, W. (2023). Campus Bridging in the Age of COVID-19. In Practice and Experience in Advanced Research Computing (pp. 355-358).**
8. Luettgau, J., Scorzelli, G., Pascucci, V., Tarcea, G., Kirkpatrick, C. R., & Taufer, M. (2022, December). NSDF-Catalog: Lightweight Indexing Service for Democratizing Data Delivery. In 2022 IEEE/ACM 15th International Conference on Utility and Cloud Computing (UCC) (pp. 1-10). IEEE.

9. Khoury, J., Klisura, D., Zanddizari, H., Parra, G. D. L. T., Rad, P., & Bou-Harb, E. Jbeil: Temporal Graph-Based Inductive Learning to Infer Lateral Movement in Evolving Enterprise Networks.
10. Lu, Y., Zhou, J., McDorman, S. T., Zhang, C., Scott, D., Bukuts, J., ... & Wang, S. (2022). Snowvision: Segmenting, identifying, and discovering stamped curve patterns from fragments of pottery. International Journal of Computer Vision, 130(11), 2707-2732.
11. Cai, G., Yu, X., Youn, C., Zhou, J., & Xiao, F. (2022). SCANNER: a web platform for annotation, visualization and sharing of single cell RNA-seq data. Database, 2022, baab086.
12. Hanfei Yu, Christian Fontenot, Hao Wang, Jian Li, Xu Yuan, and Seung-Jong Park. 2023. Libra: Harvesting Idle Resources Safely and Timely in Serverless Clusters. In Proceedings of the 32nd International Symposium on High-Performance Parallel and Distributed Computing (HPDC '23). Association for Computing Machinery, New York, NY, USA, 181–194. https://doi.org/10.1145/3588195.3592996
13. Chen, R., Li, F., Bieger, D., Song, F., Liang, Y., Luna, D., ... & Pamidighantam, S. (2022, October). CyberWater: An Open Framework for Data and Model Integration in Water Science and Engineering. In Proceedings of the 31st ACM International Conference on Information & Knowledge Management (pp. 4833-4837).
14. Schneider, B., Bartschat, K., Hamilton, K., Carr, L., Bray, I., Scrinzi, A., ... & Fischer, C. (2022). A Science Gateway for Atomic, Molecular and Optical Science (AMOS): Democratizing AMOS Research and Education. In APS Division of Atomic, Molecular and Optical Physics Meeting Abstracts (Vol. 2022, pp. U04-010).
15. Brookes, E., Rocco, M., Vachette, P., & Trewhella, J. (2023). AlphaFold-predicted protein structures and small-angle X-ray scattering: insights from an extended examination of selected data in the Small-Angle Scattering Biological Data Bank. Journal of Applied Crystallography, 56(4).
16. Brookes, E. H., & Rocco, M. (2023). Beyond the US-SOMO-AF database: a new website for hydrodynamic, structural, and circular dichroism calculations on user-supplied structures. European Biophysics Journal, 1-8.
17. Brookes, E., & Rocco, M. (2022). A database of calculated solution parameters for the AlphaFold predicted protein structures. Scientific reports, 12(1), 7349.
18. Trewhella, J., Vachette, P., Bierma, J., Blanchet, C., Brookes, E., Chakravarthy, S., ... & Zuo, X. (2022). A round-robin approach provides a detailed assessment of biomolecular small-angle scattering data reproducibility and yields consensus curves for benchmarking. Acta Crystallographica Section D: Structural Biology, 78(11).
19. E. H. Brookes, M. Rocco. US-SOMO-AF: a database of hydrodynamic, circular dichroism, and SAXS-derived parameters for the AlphaFold-predicted protein structures. 12 September 2022. International Small-Angle Scattering Conference, Campinas, São Paulo, Brazil.
20. E. H. Brookes, M. Rocco. US-SOMO-AF: a database of hydrodynamic, circular dichroism, and SAXS-derived parameters for the AlphaFold-predicted protein structures. 14 July 2022. The 25th International Analytical Ultracentrifugation Symposium. Lethbridge, CA.

21. Porter, S., Bryans, M., Vemu, S., Kamajaya, A., Ives, F., Dillman, A., ... & Smith, T. (2023). Igniting Creativity: Hackathons for Developing Undergraduate Research Projects in Antibody Engineering. Journal of advanced technological education, 2(2).
22. Diana, B., Ash, M., & Boyce, J. K. (2023). Just decarbonization? Environmental inequality, air quality, and the clean energy transition. Industrial and Corporate Change, 32(2), 304-316.
23. Boyce, J. K., Ash, M., & Ranalli, B. (2023). Environmental justice and carbon pricing: can they be reconciled?. Global Challenges, 7(4), 2200204.
24. Caton, S., Baughman, M., Haas, C., Chard, R., Foster, I., & Chard, K. (2022, November). Assessing the Current State of AWS Spot Market Forecastability. In 2022 IEEE/ACM International Workshop on Interoperability of Supercomputing and Cloud Technologies (SuperCompCloud) (pp. 8-15). IEEE.
25. Whalen, L., Bickel, N., Comandur, S., Craven, D., Dubinsky, S., & Valafar, H. Wordification: A New Way of Teaching English Spelling patterns.
26. Thomas, R. Q., McClure, R. P., Moore, T. N., Woelmer, W. M., Boettiger, C., Figueiredo, R. J., ... & Carey, C. C. (2023). Near‐term forecasts of NEON lakes reveal gradients of environmental predictability across the US. Frontiers in Ecology and the Environment, 21(5), 220-226.
27. Smith, J.W., Thomas, R.Q. & Johnson, L.R. Parameterizing Lognormal state space models using moment matching. Environ Ecol Stat 30, 385–419 (2023). https://doi.org/10.1007/s10651-023-00570-x
28. Smith, J.W., Johnson, L.R. & Thomas, R.Q. Assessing Ecosystem State Space Models: Identifiability and Estimation. JABES 28, 442–465 (2023). https://doi.org/10.1007/s13253-023-00531-8
29. Lofton, M. E., Howard, D. W., Thomas, R. Q., & Carey, C. C. (2023). Progress and opportunities in advancing near-term forecasting of freshwater quality. Global Change Biology, 29(7), 1691–1714. doi:10.1111/gcb.16590
30. Hounshell, A. G., D'Acunha, B. M., Breef-Pilz, A., Johnson, M. S., Thomas, R. Q., & Carey, C. C. (2023). Eddy covariance data reveal that a small freshwater reservoir emits a substantial amount of carbon dioxide and methane. Journal of Geophysical Research: Biogeosciences, 128, e2022JG007091. https://doi.org/10.1029/2022JG007091
31. Clark, C.M., Thomas, R.Q. & Horn, K.J. Above-ground tree carbon storage in response to nitrogen deposition in the U.S. is heterogeneous and may have weakened. Commun Earth Environ 4, 35 (2023). https://doi.org/10.1038/s43247-023-00677-w
32. Clark, C. M., Phelan, J., Ash, J., Buckley, J., Cajka, J., Horn, K., … Sabo, R. D. (2023). Future climate change effects on US forest composition may offset benefits of reduced atmospheric deposition of N and S. Global Change Biology, 29(17), 4793–4810. doi:10.1111/gcb.16817
33. Thomas, R. Q., Boettiger, C., Carey, C. C., Dietze, M. C., Johnson, L. R., Kenney, M. A., … Contributors, C. (2023). The NEON Ecological Forecasting Challenge. Frontiers in Ecology and the Environment, 21(3), 112–113. doi:10.1002/fee.2616

34. Hiltemann, S., Rasche, H., Gladman, S., Hotz, H. R., Larivière, D., Blankenberg, D., ... & Batut, B. (2023). Galaxy Training: A powerful framework for teaching!. PLoS computational biology, 19(1), e1010752.
35. Nagampalli VijayKrishna, Jayadev Joshi, Nate Coraor, Jennifer Hillman-Jackson, Dave Bouvier, Marius van den Beek, Ignacio Eguinoa, Frederik Coppens, John Davis, Michał Stolarczyk, Nathan C Sheffield, Simon Gladman, Gianmauro Cuccuru, Björn Grüning, Nicola Soranzo, Helena Rasche, Bradley W Langhorst, Matthias Bernt, Dan Fornika, David Anderson de Lima Morais, Michel Barrette, Peter van Heusden, Mauro Petrillo, Antonio Puertas-Gallardo, Alex Patak, Hans-Rudolf Hotz, Daniel Blankenberg, Expanding the Galaxy's reference data, Bioinformatics Advances, Volume 2, Issue 1, 2022, vbac030, https://doi.org/10.1093/bioadv/vbac030
36. Meyer, T. C., Carter, S., & Corbin, N. (2023). Unidata AWIPS Hosted in the Cloud . Proceedings, 39th Conference on Environmental Information Processing Technologies, 103rd AMS Annual Meeting. https://ams.confex.com/ams/103ANNUAL/meetingapp.cgi/Paper/419377
37. Weber, J., Dye, D., & Romine, P. (2023). A Network System for Environmental Monitoring, Data Sovereignty and Data Governance on Tribal Lands. 39th Conference on Environmental Information Processing Technologies, 103rd AMS Annual Meeting. https://ams.confex.com/ams/103ANNUAL/meetingapp.cgi/Paper/419480
38. Chastang, J., Corbin, N., Davis, E., Espinoza, B., & Vance, T. (2022). Unidata Science Gateway Reimagined: Unifying Access to Educational and Research Resources. Gateways 2022. https://doi.org/10.5281/zenodo.7201472
39. Espinoza, B., Chastang, J., Weber, J., Dye, D., & Romine, P. (2022). Democratizing Access to Atmospheric Modeling with WRF employing NSF Cloud Computing Resources. Gateways 2022.
40. Chastang, J. and Corbin, N. and Davis, E. and Espinoza, B. and Vance, T., Unidata Science Gateway Reimagined: Unifying Access to Educational and Research Resources, 2022, Oct.~18-20, Gateways 2022, San Diego, California, USA, 10.5281/zenodo.7201472, https://doi.org/10.5281/zenodo.7201472
41. Espinoza, B. and Chastang, J. and Weber, J. and Dye, D. and Romine, P., Democratizing Access to Atmospheric Modeling with WRF employing NSF Cloud Computing Resources, 2022, Oct.~18-20, Gateways 2022, San Diego, California, USA
42. Ramamurthy, M. and Chastang, J., The use of the Unidata Science Gateway as a cyberinfrastructure resource to facilitate education and research during COVID-19, 2022, May~23-27, EGU General Assembly 2022, Vienna, Austria, https://meetingorganizer.copernicus.org/EGU22/EGU22-10615.html
43. Zonca, A. and Chastang, J., Distributed computing on the cloud for Science Gateways with Dask, 2022, Apr.~5-7, Mini Gateways 2022, Online, 10.6084/m9.figshare.19674516.v1, https://doi.org/10.6084/m9.figshare.19653939.v1
44. Ramamurthy, M. and Chastang, J. and Espinoza, A., Unidata Science Gateway: A research infrastructure to advance research and education in the Earth System Sciences, 2023, Jun.~13-15, 15th International Workshop on Science Gateways, Tübingen, Germany.

45. Ramamurthy, M. and Chastang, J. and Espinoza, A., Unidata Science Gateway: A research infrastructure to advance research and education in the Earth System Sciences, 2023, Apr.~23-28, EGU General Assembly 2023, Vienna, Austria, 10.5194/egusphere-egu23-4682, https://doi.org/10.5194/egusphere-egu23-4682

46. Chastang, J. and Espinoza, A. and Ramamurthy, M., Unidata Science Gateway: Past, Present, and Future Plans, 2023, Jan.~8-12, Proceedings, 39th Conference on Environmental Information Processing Technologies, 103rd AMS Annual Meeting, Denver, Colorado, USA, 10.6084/m9.figshare.22043642, https://doi.org/10.6084/m9.figshare.22043642

47. Gilbert, D. G. (2022). Genes ruler for genomes, Gnodes, measures assembly accuracy in animals and plants. bioRxiv. doi:10.1101/2022.05.13.491861

48. **Michael Dodge II, Sean Cleveland, and Gwen A. Jacobs. 2022. Using Containers and Tapis to Structure Portable, Composable and Reproducible Climate Science Workflows. In Practice and Experience in Advanced Research Computing (PEARC '22). Association for Computing Machinery, New York, NY, USA, Article 76, 1–2. https://doi.org/10.1145/3491418.3535126**

49. Yick Ching Wong, Sean B. Cleveland, and Gwen A. Jacobs. 2022. Experience Migrating a Pipeline for the C-MĀIKI gateway from Tapis v2 to Tapis v3. In Practice and Experience in Advanced Research Computing (PEARC '22). Association for Computing Machinery, New York, NY, USA, Article 44, 1–4. https://doi.org/10.1145/3491418.3535139

50. Lucas, Matthew P., Ryan J. Longman, Thomas W. Giambelluca, Abby G. Frazier, Jared Mclean, Sean B. Cleveland, Yu-Fen Huang, and Jonghyun Lee. "Optimizing Automated Kriging to Improve Spatial Interpolation of Monthly Rainfall over Complex Terrain". Journal of Hydrometeorology 23.4 (2022): 561-572. doi:https://doi.org/10.1175/JHM-D-21-0171.1

Pre-prints

51. Hayashi, S., Caron, B., Heinsfeld, A. S., Vinci-Booher, S., McPherson, B. C., Bullock, D. N., ... & Pestilli, F. (2023). brainlife. io: A decentralized and open source cloud platform to support neuroscience research. arXiv preprint arXiv:2306.02183.

52. Larsen, A. H., Brookes, E., Pedersen, M. C., & Kirkensgaard, J. J. K. (2023). Shape2SAS--a web application to simulate small-angle scattering data and pair distance distributions from user-defined shapes. arXiv preprint arXiv:2301.04976.

53. Huang, X., Struck, T. J., Davey, S. W., & Gutenkunst, R. N. (2023). dadi-cli: Automated and distributed population genetic model inference from allele frequency spectra. bioRxiv, 2023-06.

54. Wannipurage, D., Deb, I., Abeysinghe, E., Pamidighantam, S., Marru, S., Pierce, M., & Frank, A. T. (2022). Experiences with managing data parallel computational workflows for High-throughput Fragment Molecular Orbital (FMO) Calculations. arXiv preprint arXiv:2201.12237.

55. Tan, W. S., Araya, E. D., Rigg, C., Hofner, P., Kurtz, S., Linz, H., & Rosero, V. (2023). Excited Hydroxyl Outflow in the High-Mass Star-Forming Region G34. 26+ 0.15. arXiv preprint arXiv:2306.07394.
56. Larsen, A. H., Brookes, E., Pedersen, M. C., & Kirkensgaard, J. J. K. (2023). Shape2SAS--a web application to simulate small-angle scattering data and pair distance distributions from user-defined shapes. arXiv preprint arXiv:2301.04976.
57. Dietze, M., R.Q. Thomas, J. Peters, C. Boettiger, A. Shiklomanov, and J. Ashander. A community convention for ecological forecasting: output files and metadata v1.0. Accepted at Ecosphere Pre-print: https://doi.org/10.32942/osf.io/9dgtq
58. D. Montgomery, H. A. Long, M. Galanko, "Utilizing Genetic and Epigenetic Data for Predicting Mutations in Colon and Pancreatic Cancers," Journal of Emerging Investigators, in press.
59. S.Y. Willow, L. Kang, D.D.L. Minh, "Learned Mappings for Targeted Free Energy Perturbation between Peptide Conformations," submitted (2023). arXiv:2306.14010
60. Measure of major contents in animal and plant genomes, using Gnodes, finds under-assemblies of several genomes. D. Gilbert, 2023, draft in progress, http://eugenes.org/EvidentialGene/other/gnodes/gnodesdoc/gnodes23daphnmodels_doc/
61. Kodama, K.M, Kourkchi, E., Longman, R.J, Lucas, M.P, Bateni, SM , Huang, YF, Kagawa-Viviani, A., Mclean, J, Cleveland. S.B, and Giambelluca T.W. Mapping daily air temperature over the Hawaiian Islands from 1990 to 2021 via an optimized piecewise linear regression technique. Journal of Geophysical Research (2023) Under Review

Other Conference Presentations / Papers

1. Awal, R.: Web-based Irrigation Scheduling Tools: IrrigWise and IrrigWise-PRISM, 2023 Water for Food Global Conference, May 08 - 11, Lincoln, Nebraska https://www.slideshare.net/waterforfood/webbased-irrigation-scheduling-tools-irrigwise-and-irrigwiseprism-new-techniques-for-irrigation-water-management-ii-2023-water-for-food-global-conferencepptx
2. Haber, Jaren, Heather A. Haveman, and Yinuo Xu. Toward computational literature reviews: Analyzing academic theories through supervised and unsupervised text-analysis methods. Paper presented at the annual meeting of the American Sociological Association, 2023.
3. Cengiz Gunay and Anca Doloc-Mihu. “AnalySim: A web platform for collaborative data sharing and analysis for research”. 32nd Annual Computational Neuroscience Meeting (CNS*2023). Lepzig, Germany. July 15-19, 2023.
4. Awal, R.: Web-based Irrigation Scheduling Tools: IrrigWise and IrrigWise-PRISM, 2023 Water for Food Global Conference, May 08 - 11, Lincoln, Nebraska https://www.slideshare.net/waterforfood/webbased-irrigation-scheduling-tools-irrigwise-and-irrigwiseprism-new-techniques-for-irrigation-water-management-ii-2023-water-for-food-global-conferencepptx

Books

None to report.

Book Chapters

None to report.

Other Products

Survey Instruments

1. **The annual Jetstream2 community survey was conducted. The full instrument and report are available at https://scholarworks.iu.edu/dspace/handle/2022/29415.**

Software or netware

1. **The software project hosted at https://exosphere.app**

Contributions from the survey:

2. Researcher software using jetstream2 as well as local resources; all is available at https://github.com/arizona-linguistics/colrc-v2
3. https://django.seagrid.org
4. https://dev.ampgateway.org
5. https://delta-topology.scigap.org
6. https://dev.delta-topology.scigap.org
7. https://geochemsim.org
8. https://gateway.futurewater.indiana.edu/
9. https://cyberwater.scigap.org/
10. https://Brainlife.io
11. Umberto Villa. (2023). hippylib/cvips_labs: Spring 2023 (Version 2023S). Zenodo. https://doi.org/10.5281/zenodo.8136705
12. Askme.lappsgrid.org
13. https://strabospot.org (StraboSpot Ecosystem)
14. https://dnasubway.cyverse.org/
15. https://earthcubeprojects-chords.github.io/chords-docs/ (EarthCube Initiative)
16. Li, Z. (., A. Michels, F. Lu, A. Padmanabhan, S. Wang (2022). CyberGIS-Jupyter for Water, HydroShare, http://www.hydroshare.org/resource/4cfd280e8eb747169b293aec2862d4f5
17. https://github.com/ryanpdwyer/pchem
18. https://js.munano.org/
19. https://http-149-165-168-30-80.proxy-js2-iu.exosphere.app/ (Utility Disconnections Data Explorer)
20. https://http-149-165-173-211-80.proxy-js2-iu.exosphere.app/ (Utility Disconnections Dashboard)
21. https://workshop.bioconductor.org
22. https://yescure.bioconductor.org

23. https://uyghur.linguistics.indiana.edu/atmo.html
24. https://analysim.tech
25. http://sig2lead.net/
26. Course materials: Umberto Villa. (2023). hippylib/cvips_labs: Spring 2023 (Version 2023S). Zenodo. https://doi.org/10.5281/zenodo.8136705
27. Hackathon materials: https://mdc2undergrad.commons.gc.cuny.edu/

Dataset

1. Corporate Toxics Information Project. 2022. "PERI US coal production GHG emissions 2020 dataset" Political Economy Research Institute, University of Massachusetts Amherst.

Theses / Dissertations

1. Kubota, A. (2023). Enabling Longitudinal Personalized Behavior Adaptation for Cognitively Assistive Robots (Doctoral dissertation, UC San Diego).
2. Klinginsmith, J. (2023). Exploring Reproducible Computing Systems Using Machine Learning Applications (Doctoral dissertation, Indiana University).
3. Liu, J. (2022). Develop the Disease Specific Bioinformatics Platforms with Integrated Bioinformatics Data. Indiana University-Purdue University Indianapolis.

Other publications

None to report.

3. Participants

3.1. Individuals

PI/CoPI as listed in research.gov

First Name	Middle	Last Name	Email	Most Senior Project Role	Nearest Person Month Worked	Contribution	Affil-iation
David		Hancock	dyhancoc@iu.edu	PD/PI	4	PI	IU
Marlon		Pierce	marpierc@iu.edu	Fomer Co-PD/PI	1	Co-PI	IU
Jeremy		Fischer	jeremy@iu.edu	Co-PD/PI	8	Co-PI	IU
Nirav		Merchant	nirav@arizona.edu	Co-PD/PI	1	Co-PI	UA
Gwen		Jacobs	gwenj@hawaii.edu	Co-PD/PI	1	Co-PI	UH

Site PIs / Co-Investigators / Senior Personnel

First Name	Mdlide	Last Name	Email	Most Senior Project Role	Nearest Person Month Worked	Contribution	Affil-iation
Enis		Afgan	enis.afgan@jhu.edu	Co-Investigator	1	Developer	JHU
Brian		Beck	bbeck@tacc.utexas.edu	Co-Investigator	5	User Services and Outreach	TACC
Sean		Cleveland	seanbc@hawaii.edu	Co-Investigator	1	System Administration	UH
Jeremy		Fischer	jeremy@iu.edu	Co-Investigator	8	Co-PI and Operations	IU
John		Fonner	jfonner@tacc.utexas.edu	Co-Investigator	0	Senior Personnel	TACC
Richard		Knepper	rich.knepper@cornell.edu	Co-Investigator	1	Senior Personnel	Cornell
John	Michael	Lowe	jomlowe@iu.edu	Co-Investigator	12	System Administration	IU
Jay		Oswald	joswald1@asu.edu	Co-Investigator	0	Senior Personnel	ASU
Mohan		Ramamurthy	mohan@ucar.edu	Co-Investigator	0	Senior Personnel	UCAR
Edwin		Skidmore	edwins@arizona.edu	Co-Investigator	3	Software Engineer	UA
Winona		Snapp-Childs	wsnappch@iu.edu	Co-Investigator	0	Project Management	IU
George		Turner	turnerg@iu.edu	Co-Investigator	3	System Administration	IU

Professional Staff

First Name	Middle	Last Name	Email	Most Senior Project Role	Nearest Person Month Worked	Contribution	Affil-iation
Stephen		Bird	stebird@iu.edu	Other Professional	10	System administration	IU
Kim		Burlingame	kb269@cornell.edu	Other Professional	0	System Administration	Cornell
Loren		Cain	lcain@tacc.utexas.edu	Other Professional	1	Project Management	TACC
Illyoung		Choi	iychoi@arizona.edu	Other Professional	3	Software Engineer	UA
Marcus		Christie	machrist@iu.edu	Other Professional	1	Software Engineer	IU
Sean		Davey	sdavey@arizona.edu	Other Professional	0	Software Engineer	UA
William		Dizon	William.Dizon@asu.edu	Other Professional	1	System Administration	ASU
Vannella		Gajjala	vgajjala@tacc.utexas.edu	Other Professional	1	Project Management	TACC
Zachary		Graber	zegraber@iu.edu	Other Professional	2	User support	IU
François		Halbach	francois@tacc.utexas.edu	Other Professional	4	System Administration	TACC
Laura		Huber	lamhuber@iu.edu	Other Professional	2	User Support	IU
Douglas		Jennewein	douglas.jennewein@asu.edu	Other Professional	1	Project Management	ASU
Johnathan		Lee	johnathan.lee@asu.edu	Other Professional	2	System Administration	ASU
Steven		Lee	shl1@cornell.edu	Other Professional	2	System Administration	Cornell
Suresh		Marru	smarru@iu.edu	Other Professional	3	Software Engineer	IU
Chris		Martin	cm10@iu.edu	Other Professional	8	Software and Systems Engineering	IU
Susan		Mehringer	shm7@cornell.edu	Other Professional	1	Training and Documentation	Cornell
Resa		Reynolds	resa.reynolds@cornell.edu	Other Professional	1	System Administration	Cornell
Julian		Pistorius	jupist@iu.edu	Other Professional	1	Software and Systems Engineering	IU
David		Schanzenbach	davidls@hawaii.edu	Other Professional	4	System Administration	UH
Ian		Shaeffer	Ian.Shaeffer@asu.edu	Other Professional	1	System Administration	ASU
Sanjana		Sudarshan	ssudarsh@iu.edu	Other Professional	3	User Services and Outreach	IU

First Name	Middle	Last Name	Email	Most Senior Project Role	Nearest Person Month Worked	Contribution	Affil-iation
Mariah		Walls	mgwall@arizona.edu	Other Professional	1	UX Developer	UA
Dimuthu		Wannipurage	dwannipu@iu.edu	Other Profesional	5	Software Engineer	IU
Le Mai		Weakley	llnguyen@iu.edu	Other Professional	6	Software Engineer	IU
Aaron		Wells	wellsaar@iu.edu	Other Professional	5	Systems Administrator	IU
Julie		Wernert	jwernert@iu.edu	Other	0	Community Survey	IU
John		Xu	zhxu73@arizona.edu	Other Professional	9	Software Engineer	UA

Undergraduate Students

1st Name	Middle	Last	email	Most Senior Role	Months	Contribution to Project	Funding Support
Junuh		Kiang	jkiang@iu.edu	Undergraduate Student	1	Undergraduate student	N/A

3.2. Partner organizations

The following partner organizations and vendors have been engaged as contractors, suppliers, or providers of software through new or existing relationships that provide a direct benefit to the Jetstream2 project.

Type of Partner organization: Commercial vendor
Name: 42on part of Fairbanks B.V.
Location: Baarn, NL
Partner's contribution to the project: Contracted support
More detail on partner and contribution: 42on provides Ceph emergency support for the open source version of the software with 24/7 1 hour response time should the project need to engage external support in the event of an outage or complex upgrade.

Type of Partner organization: Commercial vendor
Name: CodeCove Solutions
Location: Tucson, AZ
Partner's contribution to the project: Contracted support and development
More detail on partner and contribution: CodeCove solutions develops new features for the Exosphere user interface for OpenStack through specific Jetstream2 improvements as well as

community developed additions. The group also provides support for Exosphere specific problems that may be reported by Jetstream2 users.

Type of Partner organization: Commercial vendor
Name: Dell, Inc.
Location: Round Rock, TX, USA
Partner's contribution to the project: Contracted support
More detail on partner and contribution: Dell manufactured and helped deploy the Jetstream2 hardware, and will provide ongoing hardware support for Jetstream2 equipment utilized in the test, development, and production systems.

Type of Partner organization: Commercial vendor
Name: GitHub, Inc.
Location: San Francisco, CA, USA
Partner's contribution to the project: In-kind support
More detail on partner and contribution: GitHub.com provides a repository for system configuration, internal documentation, and a sharing mechanism for Jetstream2 partners. These services are provided at no charge for the Jetstream2 project.

Type of Partner organization: Academic Partner
Name: Globus, University of Chicago and Argonne National Labs
Location: Chicago, IL, USA
Partner's contribution to the project: Contracted support
More detail on partner and contribution: The Jetstream2 project leverages the XSEDE Globus subscription for basic functionality and adds additional features such as high-assurance data transfers and licensing of the Ceph connector.

Type of Partner organization: Commercial vendor
Name: MathWorks, Inc
Location: Natick, MA, USA
Partner's contribution to the project: In-kind support
More detail on partner and contribution: MathWorks enables end users of the Jetstream2 system to run MATLAB™ (and any other MathWorks products the end user is licensed to use). This is accomplished through leveraging an IU site license for this capability.

Type of Partner organization: Commercial vendor
Name: NVIDIA, Inc.
Location: Santa Clara, CA, USA
Partner's contribution to the project: In-kind support and supplier
More detail on partner and contribution: NVIDIA is a primary supplier to Dell for much of the Jetstream2 equipment. They also provide software directly to the community in the form of drivers, compilers, and containers that will be used by the Jetstream2 project. IU's partnership

with NVIDIA will include joint workshops and training for end users as well as leveraging community events put on by OpenACC where Jetstream2 will be a resource.

Type of Partner organization: Commercial vendor
Name: Qualtrics, Inc.
Location: Provo, UT, USA
Partner's contribution to the project: Contracted support
More detail on partner and contribution: Qualtrics provides the survey platform used for project evaluation. The Jetstream2 project leverages an IU site license for this capability.

Type of Partner organization: Commercial vendor
Name: Read the Docs, Inc.
Location: Portland, OR, USA
Partner's contribution to the project: In-kind support
More detail on partner and contribution: Readthedocs.io provides the primary external-facing documentation for the Jetstream2 community. This is provided at no charge to academic entities with advertising. IU contributes donations to the project to remove ads.

Type of Partner organization: Commercial vendor
Name: Slack, Inc.
Location: San Francisco, CA, USA
Partner's contribution to the project: Contracted support
More detail on partner and contribution: Slack provides a messaging platform used by the Jetstream2 project as well as directly for some service operators that rely on Jetstream2 services for science gateways and advanced uses of the system.

3.3. Have other collaborators or contacts been involved?

No.

4. Impact

4.1. What is the impact on the development of the principal discipline(s) of the project?

A number of papers have been published regarding Jetstream2 as listed in the Products section. The work published by the project team has influenced research cloud systems at smaller institutions in the US as well as the mdx system in Japan. The improvements to

Exosphere through the Jetstream2 project also have wider benefits to the OpenStack community.

4.2. What is the impact on other disciplines?

The project team has impacted other disciplines through numerous science gateways hosted and attendees from multiple disciplines participating in events at *Practice* & *Experience in Advanced Research Computing* 2023 (PEARC'23), the OpenInfra Summit, and both the United States and International Supercomputing Conferences through hosted birds-of-a-feather sessions, presentations, and workshops led by team members.

4.3. What is the impact on the development of human resources?

We have started a pilot internship program with two interns employed during this reporting period. This program focuses on training students to be able to support highly interoperable CI resources with an emphasis on system administration (configuration, operation), user support, as well as innovative technologies deployment (production-quality infrastructure to enable adoption of innovative capabilities). The goal is to recruit 33% women and 33% individuals from underrepresented groups -- this goal represents significant progress for Indiana University as the demographics for the entire IT space within IU has lower representation.

4.4. What is the impact on teaching and educational experiences?

Of the resources available through the XSEDE program, Jetstream was the most heavily used system for educational purposes. We anticipate that this will continue into ACCESS with the Jetstream2 system as the number of courses and workshops on the platform is already significant during the early operations period. The courses and workshops utilizing Jetstream2 continue to grow as listed in Section 1.4. These efforts are inclusive of small colleges, universities, and well-established research centers. Jetstream2 staff also taught elements of the ACCESS Student Training and Engagement Program (STEP-1) for its inaugural cohort.

4.5. What is the impact on physical resources that form infrastructure?

Nothing to report.

4.6. What is the impact on institutional resources that form infrastructure?

Nothing to report.

4.7. What is the impact on information resources that form infrastructure?

Nothing to report.

4.8. What is the impact on technology transfer?

There has been an important transfer of information from the Jetstream2 project to major open source technology projects, particularly OpenStack. The configuration files and methods used to deploy Jetstream2 are also publicly available and have been shared for community adoption. There have been no formal invention disclosures, but Jetstream2 staff and research faculty play an important role in the development of OpenStack. Members of the Jetstream2 team are also leading collaboration efforts in the scientific community through the OpenStack Scientific Working Group. OpenStack itself is continually transferred from its open source repository to anyone in the US that wishes to download, install, and use it.

4.9. What is the impact on society beyond science and technology?

Through its impact on the OpenStack project, the team plays an important role in the ongoing improvement of the most widely used open source cloud environment for small business and industry. This aids the development of the cloud-based economy in the US generally. In addition, through student programs, the team is able to provide high-impact workforce development opportunities to students across the nation, as students who participate will take the skills and knowledge back with them to their home institutions and will be able to share and positively impact others at their home institutions.

4.10. What percentage of the award's budget was spent in a foreign country?

0%, any engagement with non-US entities did not utilize NSF funding.

5. Changes/ Problems

5.1. Changes in approach and reasons for change

Nothing to report.

5.2. Actual or Anticipated problems or delays and actions or plans to resolve them

All planned and unplanned outages for the annual reporting period are documented below. We started tracking these outages in production rather than in early operations and will continue to do so throughout the life of this project.

Entity	Start	End	Duration	Pro Rata	Planned / Unplanned	System, Services, or Components	User impact	Comments
All clouds	23.06.29 15:52ET	23.06.29 23:16ET	7.4 hrs (7h24m)	100%	Unplanned	IU Cloud	All instances shut down, all instance connectivity lost, Keystone for satellite regions lost	power outage
IU Primary	23.06.28 19:46ET	23.06.29 09:53ET	14.12 hrs (14h7m)	25%	Unplanned	IU network fabric	Some users lost DNS on their instances	
IU Primary	23.04.16 07:00ET	23.04.16 18:07ET	11.1hrs (11h7m)	75%	Planned	General-purpose (CPU) instances	All CPU instances shut down	Building cooling outage
IU Primary	23.03.08 01:10ET	23.03.08 11:40ET	10.5hrs (10h30m)	100%	Unplanned	All	partial storage and network outage, complete user interface outage	Network fabric issue
All clouds	23.02.07 10:00ET	23.02.07 12:00ET	2hrs (120m)	100%	Planned	APIs	APIs unavailable (and user interface as a result)	Planned maintenance

IU Primary	23.01.26 23:49ET	23.01.27 00:42ET	0.88hrs (53m)	100%	Unplanned	Networking	Loss of network connectivity to/from instances, Exosphere inoperative	r01c64 stopped responding on the network
IU Primary	22.11.29 21:06ET	22.11.30 09:22ET	12.27hrs (12h16m)	62%	Unplanned	Networking	Loss of routing to certain CIDR blocks while others failed over properly	Upstream router hardware failure
IU Primary	22.11.22 18:21ET	22.11.22 19:57ET	1.6hrs (1h36m)	100%	Unplanned	Network node	Degraded network connectivity to/from instances, Exosphere inoperative	r01c64 soft lockup
IU Primary	22.09.21 15:08 ET	22.09.21 16:19 ET	1.18hrs (1h11m)	100%	Unplanned	Compute nodes	All running VMs unavailable	Vendor-induced coolant issue caused thermal shutdown
All clouds	22.09.20 13:13ET	22.09.20 14:30ET	1.28hrs (1h17m)	100%	Unplanned	Galera DB cluster	Running VMs not affected, no auth or OS commands accepted	

5.3. Changes that have significant impact on expenditures

Nothing to report.

5.4. Significant changes in use or care of human subjects

Nothing to report.

5.5. Significant changes in the use or care of vertebrate animals

Nothing to report.

5.6. Significant changes in the use or care of biohazards

Nothing to report.

5.7. Has there been a change in your primary performance site location from the originally proposed? If so, please provide the location of your new primary performance site and the reason for the change in location.

Nothing to report.