FIll in sections below, then cut and paste into the report on research.gov. Due date: September 30th, 2016. Plan to submit this on September 23rd. I use Heading 1, 2, etc for the report sections. I use [TEXT GOES HERE] to indicate where you should put text. Text in red is instructions from research.gov.

Milestones: please see what we submitted to Dan K and address these in the report: https://docs.google.com/document/d/1Z7tuKSUjZCCQpwQOSIq3oURTfnHnWGFWx8-uIP139Aw/edit#heading=h.jch3hpwyme5

For outreach, see https://docs.google.com/document/d/1KsgYZqF3aLU6Vs0Xf1i2Gu9_k5wlW81hxBNgjCt3Fg8/edit#heading=h.v0qfwhj3h1bw

Accomplishments

For NSF purposes, the PI should provide accomplishments in the context of the NSF merit review criteria of intellectual merit and broader impacts, and program specific review criteria specified in the solicitation. Please include any transformative outcomes or unanticipated discoveries as part of the Accomplishment section.

What are the major goals of the project? 8000 char limit.

List the major goals of the project as stated in the approved application or as approved by the agency. If the application lists milestones/target dates for important activities or phases of the project, identify these dates and show actual completion dates or the percentage of completion.

The overall goal of the SciGaP collaboration is to develop a Science Gateway Platform as a service that can be used to power both participating gateways (UltraScan, CIPRES, NSG) as well as the gateway community at large. To meet this goal, we are implementing SciGaP as a multi-tenanted service (currently hosted on IU's Intelligent Infrastructure) that provides a well-defined Application Programming Interface (API) for client gateways. Apache Airavata software powers persistent SciGaP services, along with third party technologies such as RabbitMQ, Zookeeper, and Keycloak. Gateways can use SciGaP services through this API to outsource the generic capabilities such as managing users, task execution on supercomputers, workflow management, experiment archiving, and digital archive sharing.

Project Year 4 saw major advances to enable this platform. This is reflected in the continued increased number of supported gateways, size of XSEDE allocations, and number of supported grants; details are in the attached project metrics document. We highlight our outreach efforts to start new science gateways in Year 4. This is built on our scalable, multi-tenanted approach to gateways, which enables us to quickly spin up gateways that use SciGaP services. Most of these make use of the PHP Gateway for Airavata (PGA), a reference implementation of the SciGaP API that we initially developed to support tutorials but which has taken on a life of its own,

The following are the project's recurring and Year 4 milestones. See also supplemental information in our metrics document.

Recurring Deliverables:

- *Engage in planned outreach activities:* We provide greater detail in the appropriate report sections below. We highlight our two PEARC17 tutorials.
- *Conduct yearly gateway community user surveys*: This report was conducted for the CIPRES gateway community user base. 9,000 survey invitations were distributed, and

there were 960 respondents. 83% of respondents reported that CIPRES helped them do something that would be difficult or impossible otherwise, 92% said their work would be slowed or halted without access to these resources. 95% reported the resource benefits their research in a tangible way. Users with very modest levels of use reported CIPRES provides value; many users who consumed less than 100 core hours still planned to publish, or reported publications that were supported by CIPRES. 84% of respondents reported they did not know what a REST API is, pointing to a clear need for education on the part of the CIPRES project. In terms of new feature requests, there was a strong interest in tools for Next Generation Sequencing, access to R and Python on the command line, and better error messaging. Users also reported about 250 publications that did not cite CIPRES, but had been supported using CIPRES resources.

- *Follow 6-week release pattern for SciGaP software patches:* as in Years 2-3, we have moved to a frequent update model for SciGaP platform services and a less frequent software release model for Apache Airavata. Service releases are organized by branches and tags, with the master branch supporting production services. Repository statistics (commits, contributors, etc) are available from the Airavata GitHub mirrors https://github.com/apache/airavata/ and https://github.com/apache/airavata-php-gateway.

Year 4: SciGaP Metrics Summary. See supplemental metrics information for a more thorough discussion of metrics.

- *Recruit new SciGaP gateways (see Outreach document). Our goal is to show increased adoption beyond Year 3.* During Year 4, we added over 20 client gateways. These range from "Software as a Service" gateways that seek to simplify delivery of newly developed scientific software to campus gateways, that seek to broaden access to campus resources.
- *Support new capabilities needed by evolution of UltraScan, CIPRES, NSG.* The ability to share digital objects created by gateways to members of groups is a cross-cutting capability. Building on earlier efforts (described last year), we have reimplemented this capability, made it a production SciGaP service, and integrated it into the Workbench Framework that powers NSG and CIPRES.
- *Complete Net+ Early Adopter Phase. Sustainability Metric: Documentation of completion by Net+.* As described in our previous reports, Net+ is better at negotiating academic price reductions with established vendors such as Box than at serving as a business incubator; we need the latter service. We are leveraging the Science Gateways Community Institute's sustainability efforts to meet this milestone.

What was accomplished under these goals (you must provide information for at least one of the 4 categories below)? 8000 char limit for each of the 4 categories.

For this reporting period describe: 1) major activities; 2) specific objectives; 3) significant results, including major findings, developments, or conclusions (both positive and negative); and 4) key outcomes or other achievements. Include a discussion of stated goals not met.

1) Major activities

Project Wide: Our major activities include the centralization of SciGaP's online services, which in turn supports our ability to scale up to support many gateways simultaneously; over 25 gateways are currently using SciGaP services. We continue to add new services as well. During the current reporting period, we deployed the Sharing Service and integrated it into both the Workbench Framework (used by CIPRES and NSG) and the PGA (the SciGaP API reference implementation gateway). This work was presented at PEARC17; see associated publication linked in this report.

We devoted significant effort to upgrading the PGA, which has emerged as an important asset to the project, providing a face to SciGaP services and a starting point for many new gateways. Upgrades include the transition from WSO2 Identity Server to Keycloak as the underlying implementation of SciGaP security services. Our primary motivation for this was to support CILogon integration in order to simplify authentication for campus gateways. We also extended SciGaP services so that they will work directly with users' local accounts on clusters. This allows SciGaP services to be more easily integrated with campus computing resources that do not support XSEDE-style community allocations.

Education efforts play an important role in SciGaP. We continued our graduate independent study course in Fall 2016 (36 students) and Fall 2017 (14 students), with an "advanced" version of the course in Spring 2017 (13 students); see http://courses.airavata.org/ for details. We hired three Research Assistants from the Spring 2016 course: Gourav Shenoy, Anuj Bhandar, and Ajinkya Dhamnaskar. These assistants served as graders and mentors for the Fall 2016 and Spring 2017 courses. Students in the Spring 2017 course prototyped several advanced features for Apache Airavata. The Fall 2017 course places increased emphasis on contributions to the Apache Airavata framework. The more stringent requirements resulted in fewer students.

UltraScan Gateway: The UltraScan and IU teams worked closely together to further improve the submission mechanism to deal more effectively with the very large numbers of multi-wavelength datasets for AUC experiments. There exists an urgency to support new instruments that are currently put online. Three manufacturers have entered the market with new hardware incorporating multi-wavelength detection. Until recently, only three prototype instruments were in use, all of them located in Germany. A new class of instruments was released earlier this year by Beckman-Coulter which produces much higher resolution data, and thus higher data density. UltraScan LIMS was upgraded to work with Airavata 0.16. The UltraScan team has upgraded a self-contained UltraScan-in-a-box solution to use a CentOS 7.x virtual machine. The solution integrates LIMS, database, GUI and HPC modules on large multi-core servers, allowing users to set up small cluster deployments at their institutions. This software integrates select codes from SciGaP developments

CIPRES Science Gateway: Major activities of the CIPRES Science Gateway include providing advice, driving requirements, and coding support for Airavata development; specifically for implementation of code for sharing jobs and data in CIPRES (and NSG) by consuming the Airavata sharing service. The sharing service for CIPRES and NSG

was created in a three separate stages. First, we modified the CIPRES user interface to allow users to manage sharing of data and jobs that have been run, and to select members to share with. Concurrently, the IU Group created the service as an Airavata feature. In the third step, code was created to allow access to the Airavata service through the CIPRES/NSG interface. The sharing service is working in our test application, and was demonstrated at PEARC17 meeting. We hope to release this feature into production in the next 6 months. We are also still considering integrating the SciGaP Identity Services with CIPRES. The integration of sharing, and all other adopted SciGaP services will be included as a configurable option in the software package. The CIPRES group participated in the science gateway instructional tutorial at PEARC17.

Neuroscience Gateway: Major activities of the Neuroscience Gateway (NSG) include adding new features, installing and implementing tools and pipelines as requested by the user community, providing feedback regarding specific NSG requirements for the SciGaP project, outreach to the user community, and involving undergraduate and high school students for internship. Examples of new activities include 1) installing the CARLsim tool (PI Jeffrey Krichmar, UC Irvine)- GPU-accelerated library for simulating large-scale spiking neural network (SNN) models with a high degree of biological detail; 2) implementing a parameter sweep tool (PI Dieter Jaeger, Emory University) - a parameter sweep tool for parameter optimization of models using the GENESIS neuronal simulation software and this was implemented using Singularity container in NSG 3) making Matlab available as a tool on NSG; 3) implementing TensorFlow via Singularity container for NSG users. Specific requirements of the NSG were provided for the SciGaP project such as data sharing related requirements. Multiple workshops, and poster presentations were done as a part of the outreach, and two high school students participated as student interns during the summer of 2017. Two UCSD undergraduate students (from underrepresented student community) participated in SciGaP funded student project with the NSG team. They compared performance of applications provided via Singularity container to their native performance and this work resulted in a PEARC17 poster presentation by these two students. NSG team also did tutorials and research talks on science gateways at the Society for Neuroscience Annual meeting in San Diego in November 2016 and at the Computational Neuroscience Annual meeting at Antwerp, Belgium in July, 2017.

2) Specific Objectives

Our project's specific objectives this year were to broaden the adoption of SciGaP services. We pursued this objective through outreach efforts that targeted gateways that support local university clusters. This takes advantage Year 3's work to bring a multi-tenanted gateway hosting capability in Apache Airavata and the ability to clone PGA instances to quickly stand up new gateways. This objective also required us to enhance SciGaP's capabilities to better integrate with campus resources. These enhancements include integration with CILogon services to support university authentication systems (many of which also support two-factor authentication) and the ability to support jobs that run under regular user accounts rather than community or group accounts. The latter allow us to more easily accommodate campus resource usage policies.

3) Significant results, including major findings, developments or conclusions

Project Wide: During this period we were able to significantly increased our support for additional gateways beyond the original three gateways, thus demonstrating the technical viability of our approach as well as the market for SciGaP-style science gateways.

Apache Airavata and the PGA: The central SciGaP concept of hosted services was amply demonstrated in Year 4, with over 25 gateways simultaneously using SciGaP services. Core improvements to Airavata included the upgrading of group management capabilities and security infrastructure.

As a significant result from the last four years of work, we see that our basic approach has succeeded, but this success is leading to two new challenges. First, we need to make significant upgrades to our reference implementation gateway, which needs to be more flexible to support the numerous gateway clients that we have recruited. These upgrades will help us continue to our current growth curve. Second, we see the need to re-implement our core services to take better advantage of container management systems such as Kubernetes, OpenWhisk, and Apache Helix. This will allow us to simplify the process of upgrading existing services and adding new services and capabilities to our core.

UltraScan Gateway: The integration of new data analysis and data acquisition methods is necessitated by new instruments available to the AUC community. Multiple methods were added to UltraScan to facilitate data import from the new instruments into a more tightly integrated LIMS experience. Users no longer have to transport files from a data acquisition computer to the analysis platform; the latest version of UltraScan imports such data directly from the instrument into the LIMS system. This reduces user error, is considerably faster, and deals efficiently with the much larger data density from the new AUC instruments. As more sites are converting from the older Proteomelab instruments to the new Optima AUC instruments from Beckman-Coulter, we have seen a rapid increase in the number of LIMS instances and the number of users and data analysis needs. Smooth operation within the UltraScan Science Gateway of these new instruments will lead to much enhanced user experience and improve data analysis performance.

NeuroScience Gateway: Since data sharing and access are of strong interest for the neuroscience community we are providing input to SciGaP data sharing and information services architecture. Based on CIPRES General Framework code NSG implemented a RESTful programmatic access (called NSG-R) for neuroscience community projects and users. A research team (PI Michele Migliore, National Research Council, Italy) as a part of the Brain Simulation Platform (BSP) project of the European Human Brain Project (HBP) utilized the Umbrella user functionality of NSG-R (RESTful service of NSG) to submit jobs to the Comet machine. The Umbrella user option allows BSP users to use a single userID from the BSP to submit jobs to HPC resources via NSG-R where individual users of BSP do not need to have an account on the NSG. NSG is being used more and more for teaching and education - example of these include NSG being used as part of the NEURON summer course, the NIH funded computational neuroscience training at University of Missouri and the NSF funded

CyberTraining project for Data-Intensive Neuroscience Learning and Research.

CIPRES Science Gateway: The CIPRES project continues to support web applications that are significant REST consumers. We have implemented several new codes - BWA, Picard, SAM tools, the Process RadTags feature of STACKS, and Bali-Phy as RESTful services for the TBAS gateway. We collaborated with the COSMIC2 gateway to create and interface and gateway for the Relion software. The distributable software underlying the CIPRES Gateway is now being implemented by two additional Gateways, the COSMIC2 gateway for cryo-electron microscopy, and another one for specialized genomics of immune development. The latter gateway has implemented the resumable HTTP upload created by the CIPRES project.

4) Key outcomes or other achievements.

Our key outcomes and achievements are described above.

Other notable achievements: Co-PI Demeler was promoted to full professor during the reporting year. Demeler has also been awarded a Fulbright Scholarship to teach UltraScan/Science Gateways abroad. The purpose of this scholarship is to bring new users to the Science Gateway from foreign countries, and build up AUC expertise with our foreign user base. Visits to Australia, Germany, UK, India, China, Scandinavia, and New Zealand are planned.

CIPRES received its 1 millionth job submission and supported its 3,500th publication in early 2017. We expect to support 1,000+ publications in 2017, and total publications will exceed 4,000 by years end.

NSG user base continues to become more diverse, compared to the neuroscience researchers that were interested primarily in spiking neuronal simulation tools since its inception in 2013. In recent years, neuroscientists interested in brain image processing, cognitive neuroscientists and neuroscientists who are involved in machine learning are using NSG. As a result software such as Matlab, R, Tensorflow etc. are requested by the users and are made available on NSG. During the 2016-2017 period, a python based neuroscience optimization tool from the European Human Brain project, a parameter sweep tool from researchers at Emory University, and a spiking neural network simulation tool for GPUs from U. California Irvine, and a NEURON/Python-based modularized framework for network simulations were made available on NSG. The parameter sweep tool was implemented via Singularity container on Comet.

The success of the Indiana University Science Gateway Group under the SciGaP and other awards has resulted in its promotion to “center” status within the Pervasive Technology Institute at IU. Pierce and Marru serve as director and deputy director, respectively.

What opportunities for training and professional development has the project provided? 8000 char limit.

Describe opportunities for training and professional development provided to anyone who worked on the project or anyone who was involved in the activities supported by the project. "Training" activities are those in which individuals with advanced professional skills and experience assist others in attaining greater proficiency.

If the research is not intended to provide training and professional development opportunities or there is nothing significant to report during this reporting period, please check "Nothing to Report" if applicable.

Project Wide: We organized two team-wide PEARC17 tutorials and conducted several

additional local efforts described below.

Apache Airavata: The IU team taught Fall 2016 (36 students), Spring 2017 (13 students), and Fall 2017 (14 students) graduate-level course in science gateway technologies. We supervised two Google Summer of Code students. Airavata developer Shameera Yodage completed his Master's Degree from IU in Fall 2016 after working as a Research Assistant for the last two years. Three IU students who took the original Spring 2016 course worked as research assistants for Spring 2017 semester before completing their Masters' degrees; they worked as hourly employees for Fall 2016. We currently have one student (Sneha Tilak) working as a research assistant and one additional hourly student in line to work as a research assistant in Spring 2018, continuing the pipeline we have established. Supun Nakandala from the IU team, as planned, as used his opportunity working on the SciGaP project to provide a foundation for his Ph. D. studies at the University of California, San Diego. We are in the process of hiring his replacement, an alumnus of our Google Summer of Code involvement. Two SUNY-Binghamton Ph. D. students traveled to SC17 and to IU to work with the IU team on Apache Airavata, with a specific emphasis on log analysis.

UltraScan Gateway: A class (BIOC5085 "Hydrodynamic Methods") was taught at UTHSCSA, which is open to students from other institutions, and was attended by investigators from Eli Lilly, University of Michigan, Colorado State University, and University of Colorado. We also taught several workshops at the 23rd international AUC conference which was held at the University of Glasgow in Scotland, UK, and presented a course on the use of analytical ultracentrifugation as a method to study macromolecular interactions at Mayo Clinic, the University of Minnesota, and at Hormel Institute for Cryo-electron microscopy. In addition, we presented multiple lectures in the US and Europe on multi-wavelength AUC and the use of the UltraScan Science Gateway as a means to analyze hydrodynamic data. Another opportunity for training was realized through multiple Google Hangout sessions, where we helped new users in Europe, India and Australia to become familiarized with the Science Gateway by being able to provide customized training to multiple users not in the US.

CIPRES Science Gateway: The CIPRES project employed 1 undergraduate student during the past year. The student was a female, with a dual major in Marine Biology and Computer Science. She worked with lead developer Terri Schwartz, PI Mark Miller, and developer Andrea Zonca to develop and implement CIPRES tools in the Jupyter Notebook environment. We presented CIPRES and NSG as part of a Gateway tutorial at the PEARC2017 conference.

Neuroscience Gateway: We have presented the NSG at multiple workshops to train users about how to use the gateway and also to collect feedback from users regarding features that users wanted for the future. NSG workshop at the Society for Neuroscience (SFN) annual meeting In November 2016 in San Diego. This NSG workshop was a half day workshop and included presentation and hands-on session on NSG as well as talks by some of the NSG users and neuroscience code developers. About 30 attendees attended the workshop. We hosted a NSG workshop in July, 2017 at the Computational Neuroscience Annual meeting in Antwerp, Belgium. This workshop was attended by about 25 people and seven talks were given by computational neuroscientists from USA and Europe. As a part of SDSC's Research Experience for High School (REHS) students program, two high school students did internship

at SDSC during the summer of 2017. Two undergraduate students also participated as interns during the 2016-2017 school year, and worked cn comparing performance of containerized versus non-containerized applications on HPC resources.

How have the results been disseminated to communities of interest? 8000 char limit.

Describe how the results have been disseminated to communities of interest. Include any outreach activities that have been undertaken to reach members of communities who are not usually aware of these research activities, for the purpose of enhancing public understanding and increasing interest in learning and careers in science, technology, and the humanities.

We engaged in the following outreach activities

- We presented two team-wide tutorials at PEARC17. The first targeted science gateway developers and the second targeted science gateway users.
- Dr. Demeler was awarded a Fulbright Scholarship to teach UltraScan/Science Gateways abroad. The purpose of this scholarship is to bring new users to the Science Gateway from foreign countries, and build up AUC expertise with our foreign user base. Visits to Australia, Germany, UK, India, China, Scandinavia, and New Zealand are planned.
- Co-PI Demeler gave the following talks and workshops: 9/16: Invited talk and AUC Seminar: University of Montana; 10/16: UltraScan Workshop: Mayo Clinic, Rochester Minnesota; 10/16: UltraScan Workshop: University of Minnesota; 10/16: Invited talk: Hormel Institute, Hormel Minnesota Electron Microscopy Center; 3/17: Invited Keynote Lecture: ARBRE-MOBIEU International Biophysics Conference, Portugal; 6/17: Invited AUC talk: University of Waterloo, Canada; 7/17: 23rd International AUC Conference in Glasgow, UK: UltraScan Workshop and Symposium featuring invited talk; 8/17: Invited AUC talk: University of Lethbridge, Canada; 8/17: Invited AUC talk: University of Montana; 9/17: Invited AUC talk: University of Western Ontario, Canada
- Supun Nakandala gave an E-Science 2016 “Hot Topic” presentation, “Apache Airavata API Security: Exploring Identity and Access Management Solutions for Multi-Tenanted eScience Framework”. Full citation is Nakandala, Supun, Hasini Gunasinghe, Suresh Marru, and Marlon Pierce. "Apache Airavata security manager: Authentication and authorization implementations for a multi-tenant escience framework." In e-Science (e-Science), 2016 IEEE 12th International Conference on, pp. 287-292. IEEE, 2016.
- Marlon Pierce gave an Apache Airavata tutorial at Gateways 2016 (November 2nd) attended by approximately 50 attendees. In addition, the Apache Airavata team gave Gateways 2016 Demo, “Co-Scheduling HPC and BigData jobs using Apache Mesos”, https://figshare.com/articles/Demo_Co-Scheduling_HPC_and_BigData_jobs_using_Apache_Mesos/4494455, a Gateways 2016 presentation, “MultiCloud Resource Management using Apache Mesos with Apache Airavata”, https://figshare.com/articles/MultiCloud_Resource_Management_using_Apache_Mesos_with_Apache_Airavata/4491629, and a Gateways 2016 presentation, “Better Data Discoverability in Science Gateways”, https://figshare.com/articles/Better_Data_Discoverability_in_Science_Gateways/4490723

- The CIPRES team gave the following presentations: “The CIPRES Science Gateway: Enabling High-Impact Science for Phylogenetics Researchers with Limited Resources” invited presentation at the DellXL Meeting in La Jolla, September 28, 2016; “The CIPRES Science Gateway after seven years. Lessons learned and best practices” at the Gateways 2016 conference in La Jolla. November 3, 2016; “The CIPRES Science Gateway, a Community Resource for Computing Large Phylogenetic Trees” Invited presentation at the DataWest Conference in La Jolla, December 15, 2016.; “A Brief Overview of CIPRES Job Management” at the Interactive Best Practices: Job Management & Scheduling webinar hosted by the Science Gateway Community Institute, August 9, 2017; “The CIPRES Science Gateway: Sustainability model” Invited presentation at the Research Data Alliance meeting (RDA2017) in Montreal CA, Sept 20, 2017; Mark Miller also served as a panelist/speaker in the session on “Best Practices in Science Gateway Job Management” at PEARC17 in New Orleans July 9, 2017.
- The NSG team gave the following presentations: Majumdar, A., Sivagnanam, S., Yoshimoto, K., Carnevale, N. T., Quintana, A., Gleeson, P. and Silver, R. A., "NSG-R Programmatic access to neuroscience applications", poster, Workshop Collaborative Development of Data-Driven Models of Neural Systems, HHMI Janelia Research Campus, Virginia, USA, Sept 18-21, 2016; S. Sivagnanam, A. Majumdar, P. Kumbhar, M. Hines, K. Yoshomoto, T. Carnevale, "Neuroscience Gateway - Understanding the scaling behavior of NEURON application, " Poster, SC16, Salt Lake City, Utah, November, 2016; N. T. Carnevale, P. Gleeson, R. A. Silver, A. Majumdar, S. Sivagnanam, K. Yoshimoto, "Seamless Integration of Neuroscience Models and Tools with High Performance Computing, " Poster, Society for Neuroscience Annual Meeting, San Diego, Nov 12-16, 2016; A. Majumdar, S. Sivagnanam, T. Carnevale, K. Yoshimoto, "The Neuroscience Gateway - Enabling Large-Scale Neuroscience Simulations and Data Processing Using Supercomputers, " Poster, Third Annual BRAIN Initiative PI Meeting, Bethesda, MD, Dec 12-14, 2016; A. Majumdar, “HPC Resources and Science Gateways,” University of California Davis, March, 2017; A. Majumdar, “Science Gateways – Access to HPC”, University of California Los Angeles, April, 2017; T. Carnevale, A. Majumdar, S. Sivagnanam, K. Yoshimoto, “The Neuroscience Gateway Portal – High Performance Computing for Neuroscientists,” Computational Neuroscience Annual Meeting (CNS 2017) Poster Presentation, Antwerp, Belgium, July 2017; A. Majumdar, S. Sivagnanam, T. Carnevale, “Neuroscience Gateway: Enabling Developers and Users to Utilize Open High Performance Computing Resources for Large Scale Simulations”, Workshop Computational Neuroscience Annual Meeting (CNS 2017), July 2017, Antwerp, Belgium; A. Majumdar, “Neuroscience Gateway,”, August, 2017, NEUROCOMP17, Madison, WI

- NSG poster was presented at the BRAIN PI meeting in December, 2016 and this was organized jointly by NIH, NSF, DOE, DARPA etc.
- A NSG workshop was held at the Computational Neuroscience Annual meeting in Antwerp, Belgium in July, 2017. Seven computational neuroscience researchers presented talks at this workshop and most of them had used NSG for their work.

What do you plan to do during the next reporting period to accomplish the goals? 8000 char limit

Supporting files. You may upload pdf files with images, tables, charts, or other graphics in support of this section. You may upload up to 4 pdf files with a maximum file size of 5 MB each.

During the next reporting period, we expect to significantly increase the number of SciGaP supported gateways. We plan to make significant revisions to the PGA, replacing the current PHP-based system with Python/Django. We expect this to be a much more modular approach that will allow gateway users to more easily customize the gateway to their needs. We also expect to re-implement Apache Airavata using a container management strategy that will provide “out of the box” fault tolerance and streamline our integration, testing and deployment processes. Our outreach efforts will continue to be based on strong participation at the PEARC, Gateways, and Supercomputing conferences during the next reporting year.

We are in the process of adapting to new XSEDE resources: Jetstream and Stampede2. Jetstream in particular requires interesting adaptations to Apache Airavata to use efficiently, as we must manage the lifecycle of virtual clusters through the OpenStack API. We anticipate the challenge here will be in making these services fault tolerant enough for day-to-day usage and not in the basic implementation.

The CIPRES and NSG projects will continue to advance the file/data sharing service and its implementation in the Workbench Framework software. We expect to release the sharing service into production in the next three months. A second goal is to integrate Airvata authentication tools into CIPRES during the next 12 months. We feel it would be useful to have access to more sophisticated authentication tools. In the case of NSG, this can decrease the overhead required for vetting users.

For the NSG project we plan to implement at least two new computational neuroscience tools/pipelines and make those available for the neuroscience community. One of these is the EEGLAB tool (https://sccn.ucsd.edu/eeglab/index.php) and another is the Human Neocortical Neurosolver (HNN http://hnn.brown.edu/). We are also planning to integrate closely with the Jetstream resource for appropriate jobs (such as parameter sweep etc.).

Products

For NSF purposes, the PI should include and discuss in the Product section the goals associated with data management and access and note any significant changes in them, as well as specific plans for dissemination of data, software and other digital research products. When you report any of these items, please include any available identifiers and whether and how these products can be accessed or shared.

INSTRUCTIONS - List any products resulting from the project during the reporting period.

If there is nothing to report under a particular item, please check, "Nothing to Report" if applicable.

Your Output Summary for this Reporting Period:

Publications

For NSF purposes, each category of publication should identify any associated data, software, other supplementary material and their appropriate identifiers. "Other publications, conference papers and presentations" should include other "non-reviewed" publications, conference papers, and presentations.

Technologies or techniques

Websites

1. http://scigap.org: general project information
2. http://airavata.apache.org/: Apache Airavata website
3. http://courses.airavata.org/

Other products, such as data or databases, physical collections, audio or video products, software or NetWare, models, educational aids or curricula, instruments, or equipment

1. https://github.com/SciGaP/: Project code repository.
2. https://github.com/airavata-courses: course website
3. https://github.com/apache/airavata: GitHub mirror for Apache Airavata source code.

Supporting Files

You may upload pdf files with images, tables, charts, or other graphics in support of this section. You may upload up to 4 pdf files with a maximum file size of 5 MB each.

[TEXT GOES HERE]: Probably nothing needed here.

Participants

For NSF purposes, for separately submitted and awarded collaborative proposals, the PI should report progress on his/her institution's portion of the collaborative effort only.

In each of the subsections below, note which collaborators or contacts are involved in data contribution and/or management.

If there is nothing significant to report during this reporting period, please check "Nothing to Report", if applicable.

* Required fields

What individuals have worked on the project?

Name	Most Senior Project Role	Nearest Person Month Worked
Marlon Pierce	PI	2
Borries Demeler	PI	1
Gary Gorbet	Programmer	12
Mark Miller	PI	2

Amit Majumdar	Co-PI	2
Terri Schwartz	Programmer	4
Kenneth Yoshimoto	Programmer	5
Subhashini Sivagnanam,	Programmer	1
Suresh Marru	Co-PI	6
Sudhakar Pamidighantam	Senior Staff	6
David Reagan	Programmer	2
Raminder Singh	Programmer	3
Lahiru Gunathilake	Graduate Student	6
Eroma Abeysinghe	Project Coordinator	3
Shameera Yodage	Graduate Student	6

What other organizations have been involved as partners?

All Participants: XSEDE, Science Gateways Community Institute

UltraScan Gateway: Juelich Supercomputing Center (Unicore integration in Airavata, and colocation services for LIMS server), Indian Institute of Science, Bangalore, India, LaTrobe University, Melbourne, Australia

NSG: Yale University; University College London, UK

CIPRES Science Gateway: NC State University, American Museum of Natural History, Oregon State University, J.Craig Venter Foundation, University of Wisconsin.

Have other collaborators or contacts been involved?

Nancy Wilkins-Diehr's S2I2 Science Gateways Community Institute. Apache Software Foundation. Von Welch, CTSC

Some significant collaborators or contacts within the recipient's organization may not be covered by "What people have worked on the project?" Likewise, some significant collaborators or contacts outside the recipient's organization may not be covered under "What other organizations have been involved as partners?"

Impacts

INSTRUCTIONS - This component will be used to describe ways in which the work, findings, and specific products of the

project have had an impact during this reporting period.

For NSF purposes, include, where appropriate, discussion of data resources and the acquisition of data skills. Include the emergence of new career paths, such as data scientists, or new disciplines.

If there is nothing significant to report during this reporting period, please check "Nothing to Report" if applicable.

Please make sure to read all instructions including NSF specific instructions, which can be found in the following link:

Required fields

What is the impact on the development of the principal discipline(s) of the project? *8000 char limit.*

Describe how findings, results, techniques that were developed or extended, or other products from the project made an impact or are likely to make an impact on the base of knowledge, theory, and research and/or pedagogical methods in the principal disciplinary field(s) of the project.

Project Wide: By unifying the common features of three different gateways into a core set of persistent, multi-tenanted services, SciGaP will greatly improve the ability of all its client gateways to offer reliable, sustainable services.

Apache Airavata: We continued work to position Apache Airavata as a tool for undertaking core cyberinfrastructure and distributed computing research as well as an operational system; this is reflected in the IU courses offered through the School of Informatics and Computing as well as our publications, listed elsewhere in the report. We continue to champion the Apache Software Foundation's open source, open community ideals as a way for federally funded software systems to be transparent, accountable, and sustainable.

UltraScan Gateway: In addition to the field of computer science, our contributions benefit the field of hydrodynamics, especially investigators in biomedical applications, material science, nanotechnology and structural, functional, and molecular biology. The new methods available now through the UltraScan gateway provide unsurpassed detail in the analysis by leveraging high-performance computing through the gateway infrastructure. These improvements are particularly important to the data analysis of data from the new multiwavelength detector, because of the sheer data density that cannot be handled without parallel processing. Additional analysis methods investigate particle size and shape distributions, as well as density and hydrodynamic radius.

CIPRES Science Gateway: In addition to our main line contribution to computer science, the project accelerates progress in virtually every field of biology, from Virology to Phylogeography.The methods made available through the CIPRES Gateway provide users around the world with access to parallel codes for computationally intensive sequence alignment and tree inference problems. This access both speeds the analysis of results, and also changes the landscape of what of which problems are computationally tractable for a given laboratory. The CIPRES Gateway has made resources available to more the 1100 institutions (Universities, Museums, Botanical Gardens, Governmental Organizations, Research Institutes) in 86 countries on 6 continents, including all 31 EPSCOR states/territories.

Neuroscience Gateway: In addition to contribution in computer science, the project has allowed computational neuroscience researchers access to HPC resources via an administratively and technologically streamlined environment for uploading models, specifying

HPC job parameters, querying running job status, receiving job completion notices, and storing and retrieving output data. This has allowed researchers to model high-dimensional parameter space exploration, study models that involve stochasticity, process MRI data using software such as Freesurfer,simulate models with large neuronal networks, run parameter sweep studies, run BluePyOpt - the Blue Brain Python Optimization Library, etc. NSG went into production in early 2013 and since then it has over 560 users from about 20 different countries.

What is the impact on other disciplines? 8000 char limit

Describe how the findings, results, or techniques that were developed or improved, or other products from the project made an impact or are likely to make an impact on other disciplines.

As described previously, a science gateway is cyberinfrastructure that directly impacts science. Gateways supported by the SciGaP award have contributed to over 780 publications in the last year. This is in addition to the core cyberinfrastructure publications by our team funded by this grant, which have been cited over 190 times.

Science gateways also support multidisciplinary work, enabling scientists in one domain to use tools from another domain that the find that they need for their research. The general purpose software and infrastructure we are developing in SciGaP, while a subject of cyberinfrastructure research itself, is also being built to provide an operational system that will be the platform for new science gateways in other scientific domains in later project years.

What is the impact on the development of human resources? 8000 char limit

Describe how the project made an impact or is likely to make an impact on human resource development in science, engineering, and technology.

Our approach to open source, open community development has created a pipeline of new developers for science gateways through the Science Gateway Architectures course and the Google Summer of Code program as well as enabling and acknowledging the contributions of peer cyberinfrastructure developers. SciGaP developers, by working in the open via Apache Airavata and GitHub, can readily demonstrate their skills to future employers or advisors.

All team members actively work with both graduate and undergraduate students. The IU team has taught basic science gateway concepts to over 50 students at IU during this reporting period. Two high school students participated in summer 2017 internship with the NSG team. They looked at the backend workflow of NSG where jobs are submitted to Comet via the NSG backend software. Two UCSD undergraduate students participated in internship during the 2016-2017 school year and looked at performance of containerized (Singularity) applications on HPC environments. One undergraduate student worked with the CIPRES group to develop Jupyter notebook implementations that can be used to run CIPRES jobs. The UltraScan group has reached out to high schools through institutional liaisons with local high schools and consistently recruited high school students to engage them in computational science, physics and chemistry disciplines.

What is the impact on physical resources that form infrastructure? 8000 char limit

Describe ways, if any, in which the project made an impact, or is likely to make an impact, on physical resources that form infrastructure, Including physical resources such as facilities, laboratories, or instruments.

Science gateways are the infrastructure that provide science-centric views of scientific computing infrastructure. The impacts of the CIPRES Science Gateway and the Neuroscience Gateway on bringing new users to XSEDE has been documented by the XSEDE project. The UltraScan gateway and its spinoffs have had a significant impact on the ability manage, analyze and understand data produced by scientific instruments. Our challenge in the upcoming years of SciGaP is to extend these successes to enable new gateways providing access to new resources.

What is the impact on institutional resources that form infrastructure?

Describe ways, if any, in which the project made an impact, or is likely to make an impact, on institutional resources that form infrastructure.

UltraScan Gateway: The developments for the UltraScan gateway significantly benefit core facilities and individual laboratories because they allow parallelized data analysis of multiple experiments simultaneously. This dramatically reduces analysis time. The LIMS backend significantly improves accuracy and the high-resolution detail obtained from our parallel methods allow core facilities like CAUMA to provide better service and higher information content to their investigators.

CIPRES Science Gateway: The CIPRES Gateway provides benefits to many institutions, by supporting experiments that could not easily be conducted with local institutional resources. Often the need for resources is confined to a single laboratory in an institution, and the consumption for many institutions is less than 50,000 core hours. In these cases it is not cost-effective to maintain a cluster for on site usage exclusively. The computational resources accessed through CIPRES make individual analyses faster, and allow multiple analyses to be run simultaneously. The result is a significant savings in local infrastructure and system administration costs.

Neuroscience Gateway: The NSG provides benefits to many institutions where computational neuroscience research is being pursued by graduate students and researchers dealing with complex and large neuronal models. These simulations require access to HPC resources and proper installation of optimized neuronal simulation tools on HPC resources and easy way to use the tools on HPC resources and retrieve output results. The NSG, by facilitating these, is allowing researchers from many institutions to carry out computational neuroscience research without having to make HPC resources and associated expertise available within their institutions. NSG has also become a dissemination platform for neuroscientists. Neuroscientists are making their tools, pipelines and libraries available via NSG for the broader neuroscience community. Examples of these include CARLsim, parameter sweep tool, BluePyOpt library, NetPyne (a python package to facilitate the development, parallel

simulation and analysis of biological neuronal networks using the NEURON simulator; PI William Lytton, SUNY Downstate).

What is the impact on information resources that form infrastructure?

(Blank)

What is the impact on technology transfer? *8000 char limit*

Describe ways in which the project made an impact, or is likely to make an impact, on commercial technology or public use.

SciGaP's Apache Airavata software is open source software that, like all Apache-licensed software, can be adapted to commercial use. Moreover, as an Apache Software Foundation project, Airavata is owned by the foundation rather by the SciGaP team members' home institutions. This clarifies governance issues such as ownership of contributions and control of the project. Our objectives in Year 5 will be to explore supplemental funding streams by partnering with universities and commercial companies.

What is the impact on society beyond science and technology? *8000 char limit*

Describe how results from the project made an impact, or are likely to make an impact, beyond the bounds of science, engineering, and the academic world.

By bringing open source, open governance methodologies to cyberinfrastructure development, we hope to provide better conduits between academic communities and commercial software development. By working in an open, public forums, we give students and junior programmers a chance to demonstrate their programming skills, problem solving abilities, and abilities to interact with a team. But we also are engaging in the other direction, bringing the thoughts and expertise of non-academic developers and architects to SciGaP problems through Apache Airavata's architecture mailing list.

Changes

INSTRUCTIONS -

The PI is reminded that the grantee is required to obtain prior written approval from the awarding agency grants official whenever there are significant changes in the project or its direction. See agency specific instructions for submission of these requests.

If not previously reported in writing to the agency through other mechanisms, provide the following additional information or state, **"Nothing to Report",** if applicable:

* Required fields

Notifications and Request

For more information on Grantee Notifications to and Requests for approval from the National Science Foundation, please

visit the Notifications and Requests section in FastLane or refer to Exhibit II-1 of the Award and Administration Guide (AAG).

Changes in approach and reasons for change 8000 char limit

Actual or Anticipated problems or delays and actions or plans to resolve them 8000 char limit

[TEXT GOES HERE]:

Changes that have significant impact on expenditures *8000 char limit*

[TEXT GOES HERE]:

Significant changes in use or care of human subjects 8000 char limit

[TEXT GOES HERE]:

Significant changes in use or care of vertebrate animals 8000 char limit

[TEXT GOES HERE]:

Significant changes in use or care of biohazards 8000 char limit

[TEXT GOES HERE]: