FIll in sections below, then cut and paste into the report on research.gov. Due date: September 30th, 2015. 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 project 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 hosted on a virtualized environment 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 WSO2 Identity Server. Gateways can use SciGaP through this API to outsource the generic capabilities such as managing users, task execution on supercomputers and workflow management.

SciGaP's operations and outreach goals are summarized below. We highlight that SciGaP met a major milestone in Year 2 for supporting multiwavelength analytical ultracentrifugation experiments, as reported by Gorbet et al in our publications. 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. As the authors state in the conclusion of the paper, "We predict that MWLD [multiwavelength data] of AUC [analytical ultracentrifugation] data will prove to be a very useful technique in a wide range of applications where complex mixtures with spectral diversity

need to be examined. The multiwavelength approach promises much improved resolution for the study of multicomponent assemblies and hetero-associating systems, and importantly, can resolve molecular weight ambiguities." The SciGaP collaboration enabled this major advance, which can be seen in the significant increase in the number of analyses performed reported in our metrics document.

The following are the project's recurring and Year 2 milestones.

Recurring Deliverables: See also supplemental information in our metrics document.

- *Engage in planned outreach activities:* We provide greater detail in the appropriate report sections below. We highlight our two XSEDE 15 tutorials.
- *Conduct yearly gateway community user surveys*: In conjunction with Nancy Wilkins-Diehr's Science Gateway Institute, we conducted a survey approximately 5,000 respondents and analyzed the results. This work resulted in accepted journal and conference publications.
- *Follow 6-week release pattern for SciGaP software patches:* We released Airavata 0.14 during the reporting period and are preparing the 0.15 release. Airavata 0.15 and 0.16 are actively used branches of the code that we use to run production UltraScan and PHP Gateway for Airavata services (PGA). We have found the Apache release process and the requirements for operating Airavata-based services to require some effort to reconcile. During the current reporting period, we made over 1700 commits to the Airavata and PGA code bases.

Year 2: SciGaP Initial Operations Metrics Summary. See supplemental information for a more thorough discussion of metrics.

- *UltraScan, NSG using SciGaP in operations.* UltraScan continues operations and is now synchronized with the latest Airavata releases. The development associated with NSG-Airavata integration is in progress and deployed to a test development portal. We expect to complete this effort early in Year 3.
- *Make SciGaP-in-a-Box Open Operations release:* This is an openly described collection of Virtual Machine resources and configurations that duplicates our production environment. We addressed this work via Ansible; our Ansible playbooks are available from our SciGaP GitHub site, https://github.com/SciGaP/Airavata-Ansible-Deployment. We have also collaborated with the XSEDE Campus Bridging Team to include Apache Airavata and the PGA gateway as part of its standard basic cluster packaging. We are participating with the Campus Bridging team in their upcoming outreach events.
- *UltraScan advanced requirements:* support new multiwavelength analytical ultracentrifugation equipment (larger jobs, more data). Capability Metric: peer reviewed publication on enhancements to SciGaP. This metric was met as described in the publication by Gorbet et al. It has also introduced advanced metascheduling requirements for Airavata that we expect to complete in Year 3.
- *Begin integration of CIPRES with SciGaP.* This work will leverage the same code

modifications as used by NSG. We expect our initial production phase of this work to be completed early in Year 3.

- *Complete Net+ evaluation.* We are re-evaluating this approach as Net+ is not well-suited to our needs. In our experience, Net+ (which has a very limited staff) is better at negotiating academic price reductions with established vendors such as Box than at serving as a business incubator. We need the latter functionality. We are currently preparing for a meeting with Dr. Brad Wheeler, IU's Chief Information Officer and Net+ co-founder to identify better options.

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: During the reporting year, our major activities were to implement Apache Airavata architectural features that will improve operational reliability and provide compatibility to work on cloud computing infrastructure. These requirements were based on our extensive experience supporting the UltraScan gateway in production. We also completed the CIPRES Workbench Framework-Airavata integration implementation, which will enable both NSG and CIPRES to route jobs through SciGaP services. Dr. Sudhakar Pamidighantam joined the IU Science Gateway Group, and we began planning the integration of his SEAGrid (formerly GridChem) science gateway with SciGaP.

During the current year we significantly expanded the administrative portions of the Airavata API. The PHP Gateway for Airavata (PGA) serves as the reference implementation of the API and was donated to the Apache Software Foundation during this year. We keep a persistent version of the PGA running at https://testdrive.airavata.org and use this for gateway developer training and daily testing. The PGA also serves as our administrative interface to SciGaP services, allowing us to add and update resources, applications, users and roles, and so on.

Apache Airavata middleware continued to evolve, based on team requirements and the new requirements for supporting the SEAGrid gateway. We deprecated Globus GRAM and GridFTP dependencies in our operations in favor of encapsulating simple SSH, SCP, and email based submissions, data movement, and state monitoring. Airavata software is still able to use Globus, UNICORE and other submission and file movement mechanisms, but we do not use this in operations except for connections to Juelich Supercomputing Center. This approach enabled us to more easily accommodate several new resources added this year over previous years and also increased our reliability. Our current success rate for jobs over the last 3 months

is 94%, compared to 90% reported the previous year. As each failure must be addressed and can be time-consuming to diagnose and recover from, this represents a substantial savings in staff time. As we collect more data, we will determine if the current improvements are statistically significant.

Airavata also underwent other significant security improvements. Based on our consultations with Von Welch's CTSC team (see joint publication with Randy Heiland et al), we completed integration between Airavata, the PGA, and WSO2's Identity Server, which we run persistently as a SciGaP service. The Identity Server provides us with identity management and role access mechanisms. This work was undertaken primarily by two Google Summer of Code students (Hasini Gunasinghe and Supun Nakandala), both of whom have been voted into the Apache Airavata project as committers. Supun will join the IU team in September 2015 as a full time software developer. We also implemented end-to-end TLS security with mutual authentication to support integration with the Neuroscience Gateway.

SciGaP services supported 13 different computing resources during the year, ranging from XSEDE to campus clusters to international computing resources at Juelich Supercomputing Center (JSC). 9 resources are in active use as of September 2015, with 4 decommissioned (XSEDE's Blacklight and Trestles, Juropa at JSC, and the original Alamo at UTHSCSA). Comet, Gordon, the new Alamo, Big Red2 (IU), Karst (IU), Mason (IU), and Jureca (at JSC) were all added resources. The revised Airavata architecture made these changes relatively smooth as we no longer were dependent on middleware providers (except at JSC) to complete the integration.

Apache Airavata: We incorporated lessons learned in operating the UltraScan science gateway and refining the API based on UltraScan, NSG, CIPRES, and SEAGrid requirements. The PHP Gateway for Airavata (PGA), previously developed outside of the Airavata framework, became an official contribution to the Airavata code base. The PGA has played an essential role as the reference implementation for the Airavata API as well as supporting XSEDE14 and XSEDE15 gateway tutorials.

UltraScan Gateway: The Airavata team worked closely with the UltraScan team to develop a new submission mechanism to deal more effectively with the very large numbers of multi-wavelength datasets for AUC experiments. The new PMG method (Parallel Masters Group) approach bundles multiple experiments into a single job which improves efficiencies in the submission process. UltraScan was adapted to the latest Airavata version and tested. UltraScan codes were implemented on four new supercomputers (Gordon and Comet at SDSC, Jureca at the Juelich Supercomputing Center, and Alamo at UTHSCSA). New algorithms for fitting and visualization of multiwavelength sedimentation velocity experiments were implemented. Two students joined the PI's laboratory for a 1 year internship (collaboration with high school) where they will develop codes for UltraScan related to performance enhancements of supercomputing algorithms. Furthermore, a new LIMS server was installed and taken into service at the Indian Institute of Science at Bangalore, India. This system is intended to be used by future users in India and Asia. Geographical proximity to the LIMS server will significantly

increase connection speeds for chatty database operations. They are significantly impacted by latencies in the network when long distances need to be traversed.

CIPRES Science Gateway: Major activities of the CIPRES Science Gateway include providing advice and driving requirements for Airavata development specifically for iterative implementation of testing of the tools for submitting jobs through the SciGaP infrastructure. The CIPRES project added two new codes and released a public RESTful service for programmatic submissions to the Gateway. Jobs submitted through the RESTful services will be handled by SciGaP in the next year. The software underlying the CIPRES Gateway was hardened into a distributable open source package from which generic Gateways can be created easily. The integration with SciGaP services will be included as a configurable option in the software package. CIPRES trained 4 interns from underrepresented groups, and supported/trained three other undergraduate employees, and mentored a female high school senior in use of digital infrastructure for phylogenetics.

Neuroscience Gateway: Major activities of the Neuroscience Gateway (NSG) include adding new features as requested by the user community, providing feedback regarding specific NSG requirements for the SciGaP project, outreach to the user community, and involving high school students for internship. Examples of new activities include 1) creating a complex pipeline utilizing multiple toolboxes to perform full brain model simulations for The Virtual Brain (TVB) neuroscience community, 2) providing MOOSE and FreeSurfer as new neuroscience software, 3) adding Comet as a HPC resource for NSG. Specific requirements of the NSG were provided for the SciGaP project such as enhancements to the API regarding remote command invocation capabilities. 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 2015. NSG team also did multiple tutorials and research talks on science gateways, cyberinfrastructure and HPC at the New Mexico State University which is a Minority Serving Institution.

2) Specific Objectives

Our project's specific objectives are described under "Major Goals". In summary, our objectives this year were to continue to validate a single API for all Airavata components, integrate it with the UltraScan gateway, and complete the integration of the Workbench Framework-based CIPRES and Neuroscience Gateways with Airavata. We are also working to integrate the SEAGrid science gateway with SciGaP.

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

Project Wide: During this period we significantly improved SciGaP operations (reducing job failures), greatly expanded the administrative portions of the API, successfully demonstrated multi-tenancy via the PGA (which is easily cloneable) combined with the WSO2 identity server, achieved integration with the Workbench Framework (which powers the NSG and CIPRES), and demonstrated SciGaP's ability to manage a wide range of scientific applications with various

idiosyncratic conventions. Detailed project metrics are given in a separate document. Highlights include an increased number of supported resources, an increase in supported grants, an increase in supported XSEDE allocations, and an increase in allocation usage.

Apache Airavata: Core improvements to Airavata enabled us to smoothly handle transitions of several resources and add campus computing resources with no middleware other than SSH, SCP, and email. This allowed us to more than double the number of resources we support and also increased our reliability. Improvements to our security infrastructure allow us to support multiple tenants for both Airavata and the PGA reference implementation gateway.

UltraScan Gateway: The integration of multi-wavelength (MW) optics in the UltraScan software has resulted in a major publication which is in press in Methods in Enzymology. It describes the processing, visualization and analysis of MW data using UltraScan and Airavata on supercomputers, and what can be learned from MW data. New algorithms developed in UltraScan result in major improvements in resolution and information content, and provide access to new and significant detail. By taking advantage of the additional spectral dimension in the experimental data, it is now possible to analyze much more complex experimental systems than was possible before. For example, for mixtures of interacting nucleic acids and proteins, or any mixture of macromolecules containing unique chromophores it is now possible to spectrally separate all such components and analyze each one separately. This makes experimental results much easier to interpret, and clearly separates solutes that are free from those which are complexed. For complexes, stoichiometries can now be derived from the spectral information in addition to the hydrodynamic information. New submission methods for large-scale jobs will enable users to more rapidly analyze the huge data amounts presented by MW experiments.

NeuroScience Gateway: Integration of SciGaP infrastructure with the CIPRES Workbench software package running NSG Portal has been completed. Two new classes that enable Workbench to call Airavata API functions have been implemented. This configuration has been used to submit a number of test NSG jobs through a SciGaP server to the SDSC Comet machine. Successful job submissions were accomplished with the following neuroscience codes: NEURON, PGENESIS, BRIAN, NEST, PyNN.

CIPRES Science Gateway: The CIPRES software platform was released as an open source project. CIPRES RESTful services were publicly released as part of the software package.

4) Key outcomes or other achievements.

Our key outcomes and achievements are described above. Collaborating as a team, we enabled multiwavelength analytical ultracentrifugation experiments and completed integration with the NSG gateway and SciGaP centralized infrastructure. Apache Airavata supported 5 XSEDE extended collaborative support projects. Our team contributed to 18 papers on cyberinfrastructure and science, and supported over 22 scientific publications by collaborators. Our h-index for new papers resulting directly from SciGaP funded work over the first two years of the project is 6. In collaboration with the Science Gateway Institute conceptualization award (led by Nancy Wilkins-Diehr), we completed a large scale survey of the science gateway

community, with over 5,000 respondents.

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 XSEDE15 tutorials, as described above, and have two additional tutorials at IU and IUPUI scheduled in November 2015. We supported Prof. Beth Plale's independent study course during Fall 2014 and are scheduled to run our own independent study course for Spring 2016. We supervised 8 Google Summer of Code students, including two students co-supervised by Dr. Emre Brookes to integrate the GenApp framework with Airavata (see publications) and 2 co-supervised by Dr. Sudhakar Pamidighantam to begin integration of the SEAGrid Gateway with Apache Airavata.

UltraScan Gateway: The UltraScan project trained four high school students during the summer and held numerous workshops to teach the use of the software. A class (BIOC5085 "Hydrodynamic Methods") was taught at UTHSCSA, which is open to students from other institutions. An online seminar series is held on the first Thursday of each month over Google Hangouts. This informal meeting presents opportunities for users to directly interact with the developers of the UltraScan software and allows them to get real-time analysis experience by following screen-shared examples. Online tutorials, videos and websites were developed to further help users learn this complicated software.

CIPRES Science Gateway: The CIPRES project employed 7 undergraduate students during the past year. The students were included 6 computer science/computer engineering majors, and 1 bioinformatics major. Four of the students were from the California Alliance to increase Minority Participation (CAMP) program. The CAMP students were given collaborative and individual tasks in developing RESTful applications for phylogenetics, or developing XML interfaces for phylogenetic codes, and worked with lead developer Terri Schwartz to improve their coding skills and experience. Two of the undergraduates worked with Wayne Pfeiffer, and were trained in the implementation and optimization of HPC codes. Two other undergraduates work with PI Mark Miller on developing and improving interfaces for the CIPRES project.

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 2014 in Washington D.C. 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 25 attendees attended the workshop. We presented a talk on NSG at the NEURON 2015 summer course at the University of California

San Diego in June 2014. About 12 attendees were present at this summer course. In March 2015 NSG team presented three short tutorials and two research talks at the New Mexico State University a Minority Serving Institution. About 60 students attended the three tutorials and about 20 attendees were present during each of the research talks. This provided opportunities for the NMSU students to learn about science gateways, HPC, cyberinfrastructure, XSEDE and XSEDE allocation etc. 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 2015. They worked on various projects related to the NSG involving web programming and NSG jobs running 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:

- SciGaP team members participated in XSEDE Extended Collaborative Support Services. PI Pierce took over as manager in February 2015 from Co-PI Marru. Six ECSS projects were assigned to the SciGaP team.
- SciGaP team members co-organized the Gateway Computing Environments 2014 and 2015 workshops in collaboration with Nancy Wilkins-Diehr, PI of the S2I2 conceptualization award for a Science Gateway Institute. Pierce also served as the software track chair for XSEDE15.
- The SciGaP team organized two XSEDE15 tutorials: "SciGaP Tutorial: Developing Science Gateways using Apache Airavata" and "XSEDE New User Tutorial: Using Science Gateways". Attendance numbers were 11 and 15, respectively. Attendee approval ratings were high (averaging >4 out of 5 for all questions), as gathered from the XSEDE15 organizers, and are available on request.
- Co-PI Marru gave a presentation on Apache Airavata and SciGaP at ApacheCon 2015 (http://events.linuxfoundation.org/sites/events/files/slides/ACNA15-Airavata-Overview.pdf).
- Team members gave presentations on SciGaP and science gateways at Internet2's Technology Exchange, the AGU Fall Meeting, the NSF SI2 PI meeting, and to the Notre Dame's Center for Research Computing.
- The team presented 3 posters at XSEDE15 and 1 poster at the Society for Neuroscience (2014) annual conference.
- The team gave or supported 8 presentations based on accepted conference papers.
- The team mentored 8 Google Summer of Code students who made important contributions to SciGaP security, integration with the GenApp framework, integration of SEAGrid gateway with Airavata and containerization via Docker.
- Four high school students (3 Voelcker Biomedical Research Academy students, and one

San Antonio NISD summer internship student) participated in research projects related to the UltraScan project, developing codes and analyzing samples using the Airavata submission infrastructure. One of the high school students attended XSEDE'15 to present an invited paper.

- Two high school students participated in summer 2015 internship and worked on NSG web programming and HPC tasks as a part of SDSC's Research Experience for High School (REHS) program. NSG team hosted workshops at Society for Neuroscience (SFN November, 2014) annual meeting, at NEURON workshop (June, 2015), and at Computational Neuroscience (CNS July, 2015) annual meeting. SFN and the NEURON workshop had about 20 attendees. In addition NSG poster was accepted at SFN 2014 and XSEDE15 conferences and were presented by NSG team members.
- The NSG team went to New Mexico State University (a Minority Serving Institution) and presented three short one hour introductory tutorial on science gateways and HPC and two research talks on topics of career in cyberinfrastructure and HPC research. About 60 students attended the three short tutorials. The research talks had about 20 attendees for each of the talks.
- Mark Miller participated as an invited panelist in the discussion group "Science Gateways: Helping Communities Lead the Way" at the April Internet2 conference in Washington D.C.

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 significant increases usage as NSG, CIPRES, and SEAGrid converge on using SciGaP services. We also expect to increase usage through regular training and outreach events (currently scheduled for November at IU and IUPUI). Our participation in the XSEDE science gateways program and potentially the Science Gateway Institute led by Nancy Wilkins-Diehr (currently under consideration for funding) will give us many additional opportunities to add gateways that cover well-known community applications. Our supplemental metrics document gives the indication of the size of impact we expect to report next year.

We do not expect major alterations in our originally planned Year 3 milestones.

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.

1. Gesing, Sandra, Rion Dooley, Marlon Pierce, Jens Kruger, Richard Grunzke, Sonja Herres-Pawlis, and Alexander Hoffmann. "Science gateways-leveraging modeling and simulations in HPC infrastructures via increased usability." In*High Performance Computing & Simulation (HPCS), 2015 International Conference on*, pp. 19-26. IEEE, 2015.
2. Lawrence, Katherine A., Michael Zentner, Nancy Wilkins-Diehr, Julie A. Wernert, Marlon Pierce, Suresh Marru, and Scott Michael. "Science gateways today and tomorrow: positive perspectives of nearly 5000 members of the research community." *Concurrency and Computation: Practice and Experience*(2015).
3. Brookes, Emre H., Nadeem Anjum, Joseph E. Curtis, Suresh Marru, Raminder Singh, and Marlon Pierce. "The GenApp framework integrated with Airavata for managed compute resource submissions." *Concurrency and Computation: Practice and Experience* (2015).
4. Pierce, Marlon E., Suresh Marru, Lahiru Gunathilake, Don Kushan Wijeratne, Raminder Singh, Chathuri Wimalasena, Shameera Ratnayaka, and Sudhakar Pamidighantam. "Apache Airavata: design and directions of a science gateway framework." *Concurrency and Computation: Practice and Experience* (2015).
5. Pierce, Marlon, Suresh Marru, Borries Demeler, Raminderjeet Singh, and Gary Gorbet. "The apache airavata application programming interface: overview and evaluation with the UltraScan science gateway." In *Proceedings of the 9th Gateway Computing Environments Workshop*, pp. 25-29. IEEE Press, 2014.
6. Marru, Suresh, Marlon Pierce, Sudhakar Pamidighantam, and Chathuri Wimalasena. "Apache Airavata as a Laboratory: Architecture and Case Study for Component-Based Gateway Middleware." In *Proceedings of the 1st Workshop on The Science of Cyberinfrastructure: Research, Experience, Applications and Models*, pp. 19-26. ACM, 2015.
7. Heiland, Randy, Scott Koranda, Suresh Marru, Marlon Pierce, and Von Welch. "Authentication and Authorization Considerations for a Multi-tenant Service." In*Proceedings of the 1st Workshop on The Science of Cyberinfrastructure: Research, Experience, Applications and Models*, pp. 29-35. ACM, 2015.
8. Lawrence, Katherine A., Nancy Wilkins-Diehr, Julie A. Wernert, Marlon Pierce, Michael Zentner, and Suresh Marru. "Who cares about science gateways?: a large-scale survey of community use and needs." In *Proceedings of the 9th Gateway Computing Environments Workshop*, pp. 1-4. IEEE Press, 2014.
9. Brookes, Emre H., Nadeem Anjum, Joseph E. Curtis, Suresh Marru, Raminder Singh, and Marlon Pierce. "GenApp module execution and airavata integration." In *Proceedings of the 9th Gateway Computing Environments Workshop*, pp. 9-12. IEEE Press, 2014.
10. .Nakandala, Supun, Sachith Dhanushka Withana, Dinu Kumarasiri, Hirantha Jayawardena, H. M. N. Dilum Bandara, Srinath Perera, Suresh Marru, and Sudhakar Pamidighantam. "Schema-independent scientific data cataloging framework." In *Moratuwa Engineering Research Conference (MERCon), 2015*, pp. 289-294. IEEE, 2015
11. Miller, Mark A., Terri Schwartz, Paul Hoover, Kenneth Yoshimoto, Subhashini Sivagnanam, and Amit Majumdar. "The CIPRES workbench: a flexible framework for creating science gateways." In *Proceedings of the 2015 XSEDE Conference: Scientific Advancements Enabled by Enhanced Cyberinfrastructure*, p. 39. ACM, 2015.

12. Miller, M. A., Schwartz, T., Pickett, B. E., He, S., Klem, E. B., Scheuermann, R. H., Passarotti, M., Kaufman, S., and O'Leary, M. A. (2015) A RESTful API for Access to Phylogenetic Tools via the CIPRES Science Gateway. *Evolutionary Bioinformatics* 11, 43-48.

Technologies or techniques

We made Apache Airavata releases 0.14 through the Apache Software Foundation during the reporting period. Apache Airavata 0.15 (used to support the production UltraScan instance) and 0.16 (the current actively developed branch) are in preparation for release.Inventions, patent applications, and/or licenses

Apache Airavata: Airavata software releases are licensed by the Apache Software Foundation using the Apache License, Version 2.

Websites

1. http://scigap.org: general project information
2. http://www.ultrascan.uthscsa.edu: UltraScan software project
3. http://uslims3.uthscsa.edu: UltraScan Science Gateway

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.

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	6
Gary Gorbet	Programmer	12
Mark Miller	PI	2
Amit Majumdar	Co-PI	2
Terri Schwartz	Programmer	4
Paul Hoover	Programmer	3
Kenneth Yoshimoto	Programmer	5
Subhashini Sivagnanam,	Programmer	2
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

UltraScan Gateway: Juelich Supercomputing Center (Unicore integration in Airavata, and colocation services for LIMS server)

NSG: Yale University

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

Have other collaborators or contacts been involved?

Nancy Wilkins-Diehr's S2I2 Science Gateway Institute conceptualization award. 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 was documented in several 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 was made resources available to more the 1100 institutions (Universities, Museums, Botanical Gardens, etc) in 73 countries on 6 continents, including 25 of 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, and simulate models with large neuronal networks and reproduce experimental results. NSG went into production in early 2013 and since then it has more than 250 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 in the previous section, science gateways in general 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 is already creating a pipeline of new developers for science gateways through the Google Summer of Code project 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. Two high school student interns worked with the NSG team during the summer of 2015. They contributed to web programming, running neuronal simulations on HPC resources by directly logging into HPC resources as well as by submitting jobs via the NSG portal. The CIPRES project trained 7 undergraduates in production software development techniques, and mentored a female high school senior in molecular biology/phylogenetics; she matriculated at Brown University in molecular biology in Sept 2015.

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 emerging 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. We expand on this below. 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 allowing experiments to be conducted that could not easily be conducted with local institutional resources. The resource provided by access 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.

What is the impact on information 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 information resources that form infrastructure,

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 Years 3-5 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]: