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 services 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 services through this API to outsource the generic capabilities such as managing users, task execution on supercomputers and workflow management. Project Year 3 saw major advances to enable this platform. This is reflected in the incr

eased number of supported gateways, supported XSEDE allocations, and supported grants; details are in the attached project metrics document. We are actively working to continue this growth trend in Year 4.

The following are the project's recurring and Year 3 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 XSEDE 16 tutorials.
- *Conduct yearly gateway community user surveys*: This report was conducted for the

CIPRES gateway community user base. 8,000 surveys invitations were distributed, and there were 963 respondents. 80% of respondents reported that CIPRES helped them do something that would be difficult or impossible otherwise, 90% said their work would be slowed or halted without access to these resources. 95% reported the resource benefits their research in a tangible way. Interestingly, even users with very modest levels of use reported CIPRES provides value, and often supported publications. 85% 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 significant interest in data sharing features, with 36% of users reporting this is very important, and another 46% reporting it was somewhat important. This provides motivation for providing a general data sharing capability through SciGaP services in Year 4. In addition, 65% of respondents felt that access to R and Python on the command line would be very useful in their work. Users also reported about 200 publications that did not cite CIPRES, but had been supported using CIPRES resources.

- *Follow 6-week release pattern for SciGaP software patches:* as in Year 2, 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. We also maintain a legacy 0.15 branch to support one older UltraScan client base. During the current year, we also made two official Apache Software Foundation software releases: Airavata 0.15 (December 2015, see https://github.com/apache/airavata/releases/tag/airavata-0.15) and Airavata 0.16 (July 2016, see https://github.com/apache/airavata/releases/tag/airavata-0.16). 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 3: SciGaP Metrics Summary. See supplemental metrics information for a more thorough discussion of metrics.

- *CIPRES, NSG, and UltraScan integrated with operational SciGaP services.* CIPRES is now using SciGaP Identity Services. UltraScan and SEAGrid are using SciGaP execution and experiment management services. NSG and CIPRES are working with the IU team to develop data and group management services.
- *Update SciGaP-in-a-Box based on full operational requirements of SciGaP's core constituent gateways.* We are revising our earlier packaging significantly to make use of Ansible scripts to deploy Airavata, the PGA, and all supporting services. Ths work is available in the Airavata GitHub repository. See https://github.com/apache/airavata/tree/develop/dev-tools/ansible. This is currently in the stable development branch.
- *Begin recruiting new gateways to use SciGaP:* We are actively engaged in this effort with several campus gateways. We have begun a focus on campus-level gateways, starting with the University of South Dakota and Oklahoma University; similar gateways

in collaboration with the University of Utah and Oklahoma State University are planned for the first quarter of Year 4.

- *Complete Net+ Validation Phase. Sustainability Metric: Documentation of completion by Net+:* As described in our previous report, 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 service. We are now exploring the NSF I/UCRC program as a more appropriate sustainability model; we have been in contact with Alan Sil and the Center for Cloud and Autonomic Computing (CAC) about joining as a partner site. We are also investigating industrial collaborations in partnership with IU's new engineering program, led by Prof. Geoffrey Fox.

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: A major technical advance this year was SciGaP's ability to support multi-tenanted gateways, which we combined with a simple way to create science gateway that are clones of the PHP Gateway for Airavata (PGA). This allows us to create new, hosted gateways quickly. We are exploiting this effort to create a number of campus-level gateways that we recruited at XSEDE16, including gateways for University of South Dakota, University of Oklahoma, Oklahoma State University, and University of Utah. We will be devoting significant effort to these and other campus gateways in Year 4. During Year 3 we also completely integrated the SEAGrid science gateway into Apache Airavata. This included a complete technology refresh of the desktop client, deprecating aging Java Web Start clients in favor of new JavaFX clients. In addition, we added an entirely new Web interface to SEAGrid based on the PGA; this web interface interoperates with the desktop tools. Finally, we highlight our education efforts. We offered Spring 2016 graduate level course at Indiana University on science gateway technologies. The class was taken by 13 students, five of whom are now working as hourlies and teaching assistants for the IU team. We are offering the course again this semester with 35 IU graduate students enrolled. See http://courses.airavata.org/ for course details.

In Year 3 we also brought SciGaP security and identity management services for SciGaP science gateway tenants into full production. This is based on extensive consultations with members of Von Welch's Center for Applied Cybersecurity Research team and is described in an accepted E-Science 2016 paper and "Hot Topics Invited Talk" (see http://escience-2016.idies.jhu.edu/program/hot-topics-invited-talks/). Our implementation is based on WSO2's open source Identity Service and uses OAuth2 and

OpenID-Connect. The CIPRES and SEAGrid Science Gateways are making use of these services, as are all new SciGaP tenants.

Finally, in Year 3 we began two major efforts that will add new capabilities for all the participating gateways. These were showcased at our mid-project All Hands Meeting, which took place at IU in June 1-2, 2016 and was attended by all project participants. First, we are actively developing new group management services for SciGaP based on Internet2's Grouper, which IU prototyped in collaboration with Grouper developers. We are now actively working on requirements and usage scenarios across the project. Second, we are developing middleware for automated data postprocessing. This was initially developed for the SEAGrid gateway and documented in an XSEDE16 paper. We are now expanding this to include CIPRES and NSG use cases. Both of these efforts will continue into Year 4.

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 will be coming online soon. Three manufacturers have entered the market with new hardware incorporating multi-wavelength detection, and soon we expect that UltraScan will have to support them. Up until this date only three prototype instruments were in use, all of them located in Germany. Commercially available multi-wavelength instruments are expected to arrive by the end of 2016. UltraScan LIMS was upgraded to work with Airavata 0.16. The UltraScan team has developed a self-contained UltraScan-in-a-box solution that 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 . Two students joined the PI's laboratory for a 1 year internship (collaboration with high school) where they developed codes for UltraScan related to grid optimization and spectral analysis of analytical ultracentrifugation data. Furthermore, a new LIMS server was installed and released into service at LaTrobe University in Melbourne, Australia. This system is intended to be used by future users in Australia, New Zealand, and eastern Asia. The geographical proximity to the LIMS server in the Pacific rim will significantly increase connection speeds for chatty database operations, which are significantly impacted by latencies in the network when long distances need to be traversed. Two instructional workshops for UltraScan users were presented in 2016: one in February at Tarrytown, NY and one in April at San Antonio.

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 code for submitting jobs through the SciGaP infrastructure. The software to manage submissions both through CIPRES and NSG was created and is ready for testing. Group members also participated during the June All Hands Meeting and in follow-on discussions in identifying other SciGaP services that can be adopted a la carte and integrated into the currently functioning CIPRES gateway. New services under consideration include authentication services and enabling sharing among user-specified groups within CIPRES, with the SciGaP Identity Services being

integrated with CIPRES in an initial effort. The CIPRES team also created a resumable HTTP-based file transfer service that can be run as an independent SciGaP service, which we will add to the SciGaP service stack. The integration with all adopted SciGaP services will be included as a configurable option in the software package. CIPRES trained two interns from underrepresented groups, and supported/trained one undergraduate from UCSD. The CIPRES group participated in the science gateway instructional tutorial and presented an analysis of user behavior on the CIPRES Gateway at XSEDE16.

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 the BluePyOpt pipeline from the BlueBrain project team and which is part of the Human Brain Project in EU; and 2) started to integrate Bridges as a resource for NSG. 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 2016. NSG team also did multiple tutorials and research talks on science gateways, cyberinfrastructure at the Janelia Research Campus as a part of the NEURON June 2016 summer course, NSG related tutorials at XSEDE16, and NSG tutorial at the Society for Neuroscience Annual meeting in Chicago in November 2015.

SEAGrid Science Gateway: SEAGrid went through a complete technology refresh during the project year, with both a new JavaFX desktop client and a Web interface based on the Airavata PGA. This effort was described in detail in an accepted XSEDE16 paper. SEAGrid is also receiving XSEDE Extended Collaborative Support from the Globus team to integrate Globus file transfer and data sharing; Globus and SciGaP identity services are being integrated (both use OAuth2 and related standards). This effort will be applicable to all SciGaP science gateways. SEAGrid usage and supported scientific research continued to increase in 2015-2016, as described in the project metrics. One particularly exciting development that will carry into Year 4 is SEAGrid's support for over 50 high school students at the North Carolina School of Science and Mathematics who will be using XSEDE resources as part of a computational chemistry class.

2) Specific Objectives

Our project's specific objectives are described under "Major Goals". This year we specifically targeted multi-tenanted gateway hosting in Apache Airavata, the ability to clone PGA instances to quickly stand up new gateways, increase Airavata's fault tolerance as we take on new gateway tenants, provide more gateway tenant-specific administration features in the PGA, and provide a scalable security infrastructure for managing gateway tenants and their users. We also had the objective of laying the foundation for a new class of features for gateway-specific post-processing data pipelines; middleware extensibility through Airavata messaging

middleware enables this.

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

Project Wide: During this period we significantly increased our support for additional gateways, which is reflected in increased XSEDE allocations and in grants supported.

Apache Airavata and the PGA: Core improvements to Airavata included the addition of group management capabilities and post-processing services. We made two major releases (0.15 and 0.16) of Apache Airavata through the Apache Software Foundation. Airavata team members mentored four Google Summer of Code students. Two of the students were particularly productive. Pradyut Madhavaram investigated how to use Apache Airavata's API with Python clients in a Jupyter notebook, which he demonstrated at our XSEDE16 tutorial and on the XSEDE science gateways community call. Jeff Kinnison developed frontend user interface for group management in the PGA that interact with Airavata's group management API, and also showed how Airavata could be extended through its messaging system to support application specific, real time monitoring. Recordings of both are available from https://www.xsede.org/gateways-symposium.

UltraScan Gateway: The integration of multi-wavelength (MWL) optics in the UltraScan software and the support of MWL data analysis continues to be the major focus of the UltraScan project; this has significant impacts on the development of Apache Airavata services. Besides development of new algorithms for the management, fitting and visualization of MWL data we continue development on codes that allow us to import data from three new instruments from three different manufacturers into UltraScan. Our group has now published four manuscripts featuring MWL data, with one more being submitted shortly. New algorithms developed in UltraScan to deal with MWL data 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 SciGaP submission methods for large-scale jobs will enable users to more rapidly analyze the huge data amounts presented by MWL experiments. This use case in particular motivated work on reliability within Apache Airavata and is motivating new efforts in using Apache Mesos for better internal job scheduling. This work will be an important effort in Year 4. A new NIH RO1 grant was awarded to the PI of the UltraScan project which focuses on the development and integration of MWL instruments in the AUC software contained in UltraScan.

NeuroScience Gateway: Long-term testing of NSG job submission through the SciGaP API was conducted. 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 is implementing a RESTful programmatic access (called NSG-R) for neuroscience community projects and users. The Open Source Brain (OSB) community project led by researchers from the University College London, UK is the first friendly user of NSG-R and they have been submitting OSB jobs via NSG-R to HPC resources.

CIPRES Science Gateway: The CIPRES project added several new REST consumers, including the TBAS project at North Carolina State, and a Galaxy implementation at Bioinformatics Center at the University of Wisconsin. We have implemented two new codes - BLAST and Trimmomatic as RESTful services for the TBAS gateway. 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 distributable software underlying the CIPRES Gateway is now being implemented by two additional Gateways, one for cryo-electron microscopy, and one for specialized genomics of immune development.

4) Key outcomes or other achievements.

Our key outcomes and achievements are described above. Our middleware infrastructure efforts over the next year will concentrate on making SciGaP group management and sharing capabilities a production service used by all member gateways, based our our current requirements discussions. We also will bring data processing pipelines into full production to support multiple interested gateways. This effort in particular will bring in the NASA-funded GeoGateway (geo-gateway.org) effort. Finally, we are working on integrating Apache Mesos as a SciGaP resource management service and will use its extensions (such as Marathon and Aurora) as internal schedulers. We do not expect major alterations in our originally planned Year 4 milestones.

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 XSEDE16 tutorials and conducted several additional local efforts described below.

Apache Airavata: The IU team taught a Spring 2016 course in science gateway technologies that was taken by 13 Computer Science students in the Master's Degree program, The Fall 2016 version the class is being taken by 35 students. We will work with the newly funded

Science Gateways Community Institute to broaden access to the course beyond IU, targetting particularly Minority Serving Institutions. We supervised four Google Summer of Code students as described above. Apache Airavata developer Supun Nakandala became a full-time developer at IU after two very successful Google Summer of Code projects. Airavata developer Shameera Yodage is completing his Master's Degree from IU in Fall 2016 after working as a Research Assistant for the last two years. The IU team also conducted two tutorials in November 2015 for IU and IUPUI.

UltraScan Gateway: The UltraScan project trained 10 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. We also taught several workshops at the 22nd international AUC conference which was held at LaTrobe University in Melbourne, Australia, and presented a course on the use of analytical ultracentrifugation as a method to study macromolecular interactions at Oulu University in Finland, where we demonstrated the use of the UltraScan Science Gateway, as well as several smaller workshops around the US.

CIPRES Science Gateway: The CIPRES project employed 5 undergraduate students during the past year. The students were all computer science/computer engineering majors. Two 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, and worked with lead developer Terri Schwartz to improve their coding skills and experience. One of the undergraduates worked with Wayne Pfeiffer, and was trained in the implementation and optimization of HPC codes.

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 2015 in Chicago. 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 22 attendees attended the workshop. We presented a talk on NSG at the NEURON 2016 summer course at the Janelia Research Campus in June 2015. About 25 attendees were present at this summer course. 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 2016. They worked on various projects related to the NSG such as writing scripts such that NSG tools can be tested automatically on a regular basis automatically.

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:

- The SciGaP team organized two XSEDE16 tutorials: “SciGaP Tutorial: Developing Science Gateways using Apache Airavata” and “XSEDE New User Tutorial: Using Science Gateways”. Both tutorials were fully attended in the provided meeting space; evaluations will be made available by XSEDE.
- Suresh Marru presented Apache Airavata at ApacheCon 2016 in Vancouver: Apache Airavata: A general purpose distributed task execution framework - Suresh Marru, Apache Software Foundation
- IU team members gave two SciGaP tutorials in November 2015 on the Bloomington and Indianapolis campuses.
- Team members presented two papers at the GCE15 workshop; see http://sciencegateways.org/past-events/gce15/.
- Team members presented 3 papers at XSEDE16 conference; see publications for details.
- The team mentored four Google Summer of Code students who contributed to Apache Airavata’s use in Jupyter notebooks, Airavata’s group-based sharing of computational experiments, and its integration with Apache Mesos.
- Ten high school students from local area high schools in San Antonio participated in research projects related to the UltraScan project in the summer of 2016, developing codes and analyzing samples using the Airavata submission infrastructure. Three students presented their work from Summer 2015/ Spring 2016 at the south Texas regional science fair, earning first prize and best in category for Biochemistry and Physics, one student earned second grand prize and attended the Intel international science fair in Phoenix, AZ. Borries Demeler chaired a conference session dedicated to multi-wavelength AUC analysis, Gary Gorbet presented technical innovations in UltraScan and the UltraScan Science Gateway during this session.
- Two high school students participated in summer 2016 internship and worked on NSG as a part of SDSC’s Research Experience for High School (REHS) program. NSG team hosted workshops at Society for Neuroscience (SFN November, 2015) annual meeting, and at NEURON workshop (June, 2016). SFN and the NEURON workshop had about 22 attendees. In addition NSG poster was accepted at SFN 2015, Supercomputing 2015 and XSEDE16 conferences and were presented by NSG team members.
- Amit Majumdar presented the benefit of science gateways and need for a collaborative environment for neuroscientists at the BRAIN Initiative PI’s meeting, organized by NSF and NIH, in December, 2015.
- Mark Miller made presentations at the Monash Undergraduate Research Projects Abroad (MURPA) colloquy.

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. Working with attendees at XSEDE16, we identified the University of Oklahoma, Oklahoma State University, the University of Utah, and the University of South Dakota as strong candidates for becoming SciGaP tenant gateways; we are pursuing the South Dakota and Oklahoma gateway efforts first, and have built prototypes under review by collaborators at these institutions. Georgia State University, Southern Illinois University, and Louisiana State University also expressed strong interest.

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

We made Apache Airavata releases 0.15 and 0.16 through the Apache Software Foundation during the reporting period. Apache 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://airavata.apache.org/: Apache Airavata website

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	6
Gary Gorbet	Programmer	12
Mark Miller	PI	2
Amit Majumdar	Co-PI	2
Terri Schwartz	Programmer	4
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, Science Gateways Community Institute

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

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 was made resources available to more the 1100 institutions (Universities, Museums, Botanical Gardens, etc) in 85 countries on 6 continents, including 29 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, process MRI data using software such as Freesurfer and simulate models with large neuronal networks. NSG went into production in early 2013 and since then it has close to 400 users from about 20 different countries.

SEAGrid Gateway: SEAGrid allows computational chemists, molecular scientists, and engineers and educators to make better use of computing facilities. SEAGrid has enabled over 30 scientific publications over the last two years.

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 has created 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 2016. 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 4 undergraduates in production software development techniques. The UltraScan project engaged 10 high school students for summer internships in analytical ultracentrifugation, program and computing projects and trained them various programming projects, data analysis and laboratory techniques. The IU team has taught basic science gateway concepts to 48 students at IU during this reporting period.

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 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 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 4-5 will be to explore supplemental funding streams by partnering with universities and commercial companies, and we will evaluate the NSF I/UCRC program for this; our open source policies will be very important for these efforts.

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]: