My Desktop
Prepare & Submit Proposals
New! Prepare Proposals (Limited proposal types)
Prepare Proposals in FastLane
Proposal Status
Awards & Reporting
Notifications & Requests
Project Reports
Award Functions
Manage Financials
Program Income Reporting
Grantee Cash Management Section Contacts
Administration
Lookup NSF ID

Preview of Award 1541215 - Annual Project Report

Cover |
Accomplishments |
Products |
Participants/Organizations |
Impacts |
Changes/Problems

Cover

Federal Agency and Organization Element to Which Report is Submitted:	4900
Federal Grant or Other Identifying Number Assigned by Agency:	1541215
Project Title:	CC*DNI DIBBs: Data Analysis and Management Building Blocks for Multi-Campus Cyberinfrastructure through Cloud Federation
PD/PI Name:	David A Lifka, Principal Investigator Thomas R Furlani, Co-Principal Investigator Richard Wolski, Co-Principal Investigator
Recipient Organization:	Cornell University
Project/Grant Period:	10/01/2015 - 09/30/2020
Reporting Period:	10/01/2018 - 09/30/2019
Submitting Official (if other than PD\PI):	N/A
Submission Date:	N/A
Signature of Submitting Official (signature shall be submitted in accordance with agency specific instructions)	N/A

Accomplishments

*** What are the major goals of the project?**

The major goals established at the launch of the 5-year Aristotle DIBBs project were:

- Implement a scalable and sustainable multi-institutional cyberinfrastructure cloud federation model that provides data analysis building blocks in support of multiple research disciplines requiring flexible workflows and analysis tools for large-scale data sets. Federation sites are Cornell University, University at Buffalo (UB), and University of California, Santa Barbara (UCSB).
- Support seven strategic science use cases from intentionally diverse disciplines (earth and atmospheric science, finance, chemistry, astronomy, civil engineering, genomics, and agriculture) to demonstrate the potential of a federated cloud as a campus bridging paradigm. Explore data analysis techniques and their applicability to different disciplines. Document tools, workflows, challenges, and best practices for each use case.
- Encourage and reward data analysis resource sharing with a new allocations and accounting model that provides a fair exchange mechanism for resource access between and across multiple institutions. Develop and build a new tool for cloud metrics into Open XDMoD that includes DrAFTS (Durability Agreement From Time Series) statistics to make online forecasts of future performance and allocations levels available to users.

A successful 36-month NSF project review occurred in March 2019. The review presentation is available here: https://federatedcloud.org/reports/Aristotle-Project-1541215-36Month-Review-Lifka-3.15.19.pdf. After discussing the panel's comments and suggestions, the project team set the following goals for the remainder of PY4 to the end of PY5 (September 30, 2020):

- Complete the portal with all sites at the same functionality level and fully interoperable.
- Package the capabilities so that others can build their own cloud federations (or join ours).
- Investigate pricing capabilities with RightScale, focusing on two or more science use cases, and benchmarking on Aristotle and a public cloud(s).
- Work with the XSEDE CRI to get other campuses to implement OpenStack clouds for research; Dartmouth is helping us harden our OpenStack documentation and training.
- Demo cloud bursting as an application level service (the ability to migrate to NSF and public cloud by containerizing applications is the main focus).
- Add new use cases if they are ready to use the federation and do not require in-depth consulting support; new use cases will push us in new directions (e.g., current "science as a service" use case VMs have been running over 2 years on Aristotle).

*** What was accomplished under these goals (you must provide information for at least one of the 4 categories below)?**

Major Activities:

Architect and install PY4 DIBBs cloud infrastructure/storage assets at 3 federation sites

Specific Objectives:

- Cornell purchased 16 nodes, a network switch, and 240TB storage to add to the Ceph pool. 448 cores were added to Red Cloud and a server with 4 GPUs. Many users are requesting GPUs.
- University at Buffalo (UB) purchased networking gear to refactor their OpenStack network and storage assets to double the size of their Ceph storage. GPUs are in test phase.
- UC Santa Barbara (UCSB) purchased network switches, brought their Aristotle OpenStack cloud online, and installed federated XDMoD cloud metrics.
- All 3 federation sites successfully transitioned from Eucalyptus to OpenStack production clouds and migrated users with minimal interruption.
- A total of 2,060 cores (includes campus resources) and 2400TB storage is now available on Cornell Red Cloud, UB Lake Effect Cloud, and UCSB Aristotle Cloud.
- Dartmouth joined the federation and is deploying their first OpenStack cloud. Globus single sign-on is ready, internal and external VLANs created, and a ~600TB production Ceph cluster underway. 100-200 cores will be deployed. Dartmouth will provide valuable feedback on federation documentation while

benefiting from Aristotle lessons learned and access to the Aristotle GitHub repository where all Puppet scripts, scripts for creating private networks, etc. are available. See *HPCwire*: https://www.hpcwire.com/off-the-wire/dartmouth-joins-aristotle-cloud-federation-to-explore-federated-cloud-computing-model/.

- Red Hat's TripleO is being investigated as an OpenStack deployment and management tool.

Develop portal code to allow Globus authentication by users

- All science users have accounts on all 3 clouds and can log into each with their Globus identity.

Develop Phase 3 portal content

- Completed the transition from a local to a federated accounting and allocations system.
- Completed the portal dashboard which displays usage data on a project level to science team members and automates account creation every 30 minutes.
- Updated the portal user guide, publications, use case products, and Aristotle news on a regular basis.

Implement Open XDMoD cloud data collection across the federation

- Data ingestion from all sites is now being monitored with federated Open XDMoD v. 8.1.2.

Integrate Metrics and Predictions into Open XDMoD and associated data collection for portal

- Integrated cloud metrics (# instances running/started/ended, # cores utilized, CPU hours per instance/total, and wallhours per instance/total, etc.) into Open XDMoD and made select metrics available on the portal dashboard.
- Developed a new DrAFTS 2.0 spot price prediction tool called "Aristotle AWS SpotPrice Visualizer."
- Investigating the ability of RightScale to analyze the cost differences of running Linear Algebra benchmarks, WRF, OpenMORDM, and Terraform/Ansible automated cluster creation on Red Cloud, AWS, Google, and Azure (proxies for apps that spin up MPI clusters on demand).

Migrate to NSF and public clouds

- All Cornell science use cases were containerized for portability to any cloud. Also, UB containerized app kernels in Docker (HPCC, HPCG, NAMD, NWChem, GAMESS) and ran on OpenStack.

Support 7 science use case teams with current cloud resources while implementing plans for federated cloud usage

1. **A Cloud-Based Framework for Visualization & Analysis of Big Geo Data** (Varun Chandola, UB). Chandola co-organized a panel at the AGU Meeting on the "Application of Information and Data Science Methods and Technologies to Climate Research and Energy-Water Knowledge Discovery" and displayed a poster on work deployed on the Aristotle cloud. New machine learning components are under development to further enhance webGlobe and a Jupyter Notebook extension of the webGlobe framework will allow users to analyze underlying data.
2. **Global Market Efficiency Impact** (Dominik Roesch, UB). Roesch, 4 PhD students, and outside university collaborators are running VMs that hold extensive finance data on Aristotle. Their analysis of the difference between asset prices during day and night was presented at two top finance

conferences: the American Finance Association and Jackson Hole. A paper on tick size and liquidity was accepted by the Journal of Financial Economics.

3. **Application of the WeatherResearch and Forecasting (WRF) Model for Climate-Relevant Simulations on the Cloud** (Sara C. Pryor, Cornell). Analyzed output from simulation of WRF model at 12Km over eastern N. America for 2001-2016 for the assessment of year-to-year variability in the wind resource (150TB WRF runs). Enhanced high-resolution simulations of wind farm wakes from 2 parameterizations to advance methods to optimize wind turbine arrays and maximize system-wide power production. Analysis of output from simulations on Aristotle was conducted on Jetstream. This work resulted in 6 conference presentations and 2 journal publications.
4. **Transient Detection in Radio Astronomy Search Data** (James Cordes, Cornell). Improved the Fast Radio Burst (FRB) Pipeline, a customizable scientific software pipeline for detecting single pulse candidates that may be FRB sources in radio astronomy data, by preparing to parallelize certain methods, updating the Python 3 version of PRESTO, and adding functionality for general radio transient detection beyond just FRBs. Added checkpoint/restart functionality to the pipeline to support tests on AWS spot instances as part of the RightScale investigation. Updated container and pipeline documentation.
5. **Water Resource Management Using OpenMORDM** (Patrick Reed, Cornell). Regenerated all published results for the multi-node MPI Lake Problem simulation which will serve as validation of our containerization methodology. Ran MPI between Docker containers on multiple VMs with the Lake Problem code on Aristotle and Jetstream. Tested the OpenMORDM group MPI WaterPaths software in a Docker container on single and multiple VMs. The goal is to show that WaterPaths can be used in the cloud with a performance comparable to that under HPC.
6. **Mapping Transcriptome Data to Metabolic Models of Gut Microbiota** (Angela Douglas, Cornell). Migrated a Windows VM to Linux VM in OpenStack, and wrote scripts and library functions to expedite running model simulations and choosing ideal model parameters. Using SteadyCom, investigated the scope of metabolic interactions that occur among a Drosophila gut microbial community. Identified 159 unique metabolites that are exchanged and showed that the gut microbial community is an important source of TCA cycle intermediates 2-oxoglutarate and succinate. An Aristotle VM enabled research published in mBio journal: “The Cost of Metabolic Interaction in Symbioses between Insects and Bacteria with Reduced Genomes.”
7. **Multi-Sourced Data Analytics to Improve Food Production and Security** (Kate McCurdy, Sedgwick; Elizabeth Grafton-Cardwell, UC Lindcove; Chandra Krintz, UCSB SmartFarm with Cal Poly, San Louis Obispo and Fresno State). Progress was made developing a low-power, low-cost, multi-tier IoT deployment for citrus frost prevention and for differential irrigation of almond trees. The Sedgwick Reserve’s “Where’s the Bear” project is using Pacific Research Platform’s Kubernetes and containers environment to train image recognition models in conjunction with Aristotle. A new use case called Citrus Under Protective Screening (CUPS) was initiated. CUPS is a potential remedy for citrus greening which has devastated citrus in FL and is now threatening CA. Aristotle will serve the data hosting service for CUPS. 8 new publications (including a Best Student Paper award) and 2 keynote presentations were produced.

Aristotle quarterly reports (https://federatedcloud.org/reports/) and publications (https://federatedcloud.org/about/publications.php) further detail these accomplishments.

Significant Results:

Key outcomes or
Other achievements:

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

Cross-Training & Knowledge Sharing

Expertise was shared between sites every two weeks on Aristotle team conference calls to ensure timely cross-training and knowledge sharing, and in-depth follow-up calls occurred to solve specific technology implementation issues and to share lessons learned.

Cross-site discussions and collaborations between use case scientists and Aristotle support teams facilitated the sharing of research tools and cloud computing technologies, e.g., sharing insights on Docker security and networking, Red Hat TripleO implementation, etc.

A train-the-trainer approach was used, i.e., the training of one use case team member was documented in order to facilitate the training of the entire research group.

Knowledge gained from events such as the all-day “Openshift Container Application Training Workshop” with hands-on lab held at Cornell were beneficial to the federation as a whole.

Finally, the project team is providing advice and guidance to Dartmouth College as they deploy their first OpenStack cloud (Dartmouth joined the federation as its first member). Dartmouth, in turn, will be providing the project team with feedback on Aristotle cloud implementation documentation and the federation’s technologies. This documentation will be used to educate users and share lessons learned on deploying OpenStack clouds. A U.S. campus cloud implementation service, made possible by a new XSEDE/Aristotle partnership, will be available nationwide starting in September 2019. See https://insidehpc.com/2019/03/xsede-teams-with-aristotle-cloud-federation-to-implement-clouds-on-u-s-campuses/ to learn more.

Undergraduate & Graduate Student Development

REU students made valuable contributions to Aristotle science use cases at Cornell and UCSB, and gained valuable domain-specific knowledge and first-hand experience using clouds for data analysis.

Jackie Zheng, a Cornell REU student, produced a script which sets up Google and AWS instances to run the Lake Problem on different instances. Originally, the Lake Problem repository created a Docker container cluster in a single instance. Jackie used Ansible and Terraform to configure the Docker containers to allow for multi-instance communication where each instance takes up a single Docker container. Jackie also created a Docker container which will run the scripts inside a container so that the host does not need additional dependencies. Jackie collaborated with Brandon Barker, Pete Vaillancourt, and Bennett Wineholt.

Cornell REU student Ryan Hill built a functioning neural network that classifies different types of radio frequency interference into one of 5 categories, including no RFI. It does so with up to 90% accuracy depending on the loss rate. Currently, Ryan is training over many different loss rates trying to identify patterns in the misclassifications. Depending on what is found, he will modify the network, possibly adding additional layers, then will extend it to classify images that include single pulses and multiple different types of RFI to see how it responds. The code for the neural network exists in a Jupyter Notebook and .py versions; in addition, Ryan created code to develop, use, and modify the RFI and single pulse graphs.

At UCSB, REU student Gareth George developed an AWS compatibility layer for microcontrollers to use as part of the Citrus Under Protective Screening (CUPS) project whose goal is to protect the CA citrus crop which is under threat from citrus greening disease that devasted FL citrus. AWS Lambda is a facility that is popular with the science team but restricted for use to AWS. This student (acting on his own initiative) has created a compatibility service that not only runs in Aristotle but can also run on resource restricted devices themselves. From a systems perspective, this is a significant breakthrough and we are now preparing a paper for publication on the effort.

UCSB REU student Gabriel Soule has become the primary developer and deployment engineer for our collaboration with the Edible Campus project (http://www.sustainability.ucsb.edu/ediblecampus/). This project is interested in using Chandra Krintz's SmartFarm technologies to monitor Edible Campus farm sites. Soule is currently siting solar power and long-distance Wi-Fi at the Edible Campus farm which is located on unincorporated land next to UCSB.

In July 2019, UCSB PhD candidate Nevena Golubovic, under the supervision of Aristotle use case scientists Krintz and Wolski, won the Best Paper Award at the IEEE International Congress on the Internet of Things (https://cs.ucsb.edu/news/3488).

During winter 2019, Aristotle resources supported a Computer Science class taught by Wolski on Cloud Computing, Edge Computing, and IoT (CS293B - https://sites.cs.ucsb.edu/~rich/class/cs293b-cloud/). The learning objective was to build a working prototype of either an application that uses edge and cloud computing technologies or a new edge/cloud technology itself. Thirty students accessed VMs on Aristotle.

Hundreds of statistics students continue to use a Docker container with RStudio on the UB cloud. This is an example of how the Aristotle project is impacting campuses at large (beyond the Aristotle project science use cases). Likewise, Cornell Institute for Social and Economic Research faculty and student projects continue to leverage Linux and Windows images built by Cornell's cloud team.

Cornell is using AWS AppStream to replace student labs on campus. This will reduce student lab costs (hardware, maintenance, etc.) and provide students with the ability to learn on any platform with a browser (Macs, Chromebooks, or PCs) and in any location (the classroom, library, or at home). Similarly, Dartmouth's first used case is focused on using Jupyter notebooks and the cloud for education to reduce the number of campus computer labs which most of the time aren't being used; Dartmouth is using Cornell's Red Cloud for testing.

Many of the Aristotle science use cases impact graduate and PhD student skills development. For example, at UB Aristotle use case lead Dominik Roesch taught 4 PhD students to use the OneTick Time-Series financial framework with the underlying data hosted on Aristotle; the students subsequently launch new investigations such as how the minimum price movement of all U.S. stock affects liquidity.

PhD, postdocs, and graduate students at Cornell, UCSB, and collaborating academic institutions receive training from Aristotle science team staff on cloud and edge computing technologies, For example, grad students and faculty from Cal Poly San Luis Obispo and Fresno State worked with Aristotle staff and farm consultants to instrument an almond test tree for differential irrigation. The goal is to see how much water can be saved by irrigating different sides (sunny and shade) of the root stock in proportion to its dryness. The deployment is a complete end-to-end, multi-tier IoT system that links Aristotle, edge systems, and sensors that target precision irrigation for almonds. Students also benefit from training on application containerization, library functions, modeling software, multiple VM execution, deployment automation, and algorithmic issues.

K-12

There were no K-12 activities this reporting period.

How-To Documentation & Training

How-to user guides continued to be developed on GitHub and, when ready, publicly released on the Aristotle portal. Build-your-own federation documentation is also being released on the portal as it is developed (https://federatedcloud.org/using/buildyourown.php).

We have been engaged in extensive containerization work with a focus on training scientists in each use case group so they, in turn, can train other members in their group.

*** How have the results been disseminated to communities of interest?**

Scientific Meetings, Publications, and Conferences

Science use case researchers and the Aristotle team presented results at scientific meetings where they referenced the Aristotle project and its contribution to their success. Use case researchers also published papers which acknowledged the project. The “Products” section of this report features 37 selected publications, conference presentations, and videos.

Aristotle PI and co-PIs are actively involved in professional societies and conferences such as the Practice & Experience in Advanced Research Computing (PEARC) conference, Coalition for Academic Scientific Computing (CASC) meetings, SC Conference, and NSF workshops all of which have afforded opportunities to share progress on the project. For example, at PEARC19 in July co-PI Tom Furlani served as Conference Chair and the Aristotle team was invited to make two presentations on the federation and campus research clouds. At the CASC 2018 Fall Meeting, Aristotle portal lead Susan Mehringer shared Aristotle lesson learned on a panel focused on “Federations.”

PI David Lifka has a leadership role in the eXtreme Science and Engineering Discovery Environment (XSEDE) project and keeps XSEDE management abreast of developments in cloud computing. Lifka also serves on the NSF Directorate for Computer, Information Science & Engineering (CISE) Advisory Committee. In addition, Resa Reynolds, Aristotle infrastructure lead, is president of the Dell XL HPC Consortium and Mehringer is the XSEDE training lead. Each of these roles afford opportunities to share cloud computing insights and experiences with their cyberinfrastructure colleagues.

At the November SC18 Conference, Cornell and the UB featured the Aristotle project in their exhibits and highlighted science used case progress and cloud technologies. The project team also meet with industry partners, universities, and OEMs and ISVs to brief them on the federation’s current status. In addition, Rich Knepper met with the public cloud providers to further investigate the cloud marketplace concept. At the 10th Workshop on Scientific Cloud Computing, co-PI Rich Wolski shared his analysis of changes in AWS spot market pricing.

Most recently, at the August 2019 NSF-sponsored Organizing Urban Transects for a Sustainable Transformation of Economic Partnerships across the Lower Great Lakes Workshop (OUTSTEP 2019 - https://outsteps.org/), use case scientist Varun Chandula demonstrated the first version of an Aristotle-powered platform whose goal is to provide a data and computing infrastructure to support researchers and planners working on sustainable development for the Lower Great Lakes area.

At the upcoming SC19 Conference, Cornell will highlight a new Aristotle and XSEDE partnership that will deliver OpenStack cloud implementation services onsite for interested U.S. campuses. Cornell is currently advising Dartmouth on an OpenStack deployment; they are the first university outside the funded partners to join the federation. PI Lifka meet with the Dartmouth CIO and scientific research computing team in Hanover, NH.

Aristotle Portal

The Aristotle portal provides scientists and the cyberinfrastructure community extensive information on the project, including detailed use case accomplishments/plans/products, publications, news and events, user guides and other documentation, and in depth quarterly project reports (https://federatedcloud.org/).

Communicating to a General Audience

News stories about Aristotle use case scientists broaden the public understanding of the value of IT and engineering in solving societal problems. For example, a December 2018 news story and video by Washington, DC's The Hill highlighted how UCSB smart farming will increase access to food (see https://thehill.com/hilltv/boundless/421073-computer-science-expert-says-smart-farming-will-increase-access-to-food), and a lecture series on transformative life scientists was chaired by Aristotle use case scientist Angela Douglas and was open to the public.

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

Infrastructure & Portal Plans

- Order/install/configure PY5 storage assets and cloud infrastructure at each site to enhance science use case research capabilities and to continue to build a sustainable federated cloud model.
- Improve the installations at each site to enhance research progress of the science use case teams and new use cases from Dartmouth (first institution to test the federation).
- Improve OpenStack operations and continue to share lessons learned.
- Enhance the portal dashboard which displays usage data on project level to science team members.
- Expand the Aristotle user guide and build-your-own cloud documentation with best practices. Add advanced and federation-specific topics.
- Continue to provide science use team training on how to use the federated cloud and topics such as how to build containers efficiently.
- Upload the portal template, database schema, etc. so campuses can build their own federations.

Metrics & Usage Plans

- Integrate additional cloud metrics into Open XDMoD and refine with use case researchers.
- Deploy containerization application kernels on all Aristotle resources.
- Deploy containerization kernels on public cloud providers.
- Benchmark and compare local public and private cloud infrastructure.
- Examine RightScale API and import, as feasible, public cloud accounting information into Open XDMoD.
- Create a public cloud realm in Open XDMoD.
- Investigate incorporating VM performance summaries into Open XDMoD.
- Experiment with a new method of determining price comparison using DrAFTS 2.0 and, once the prototype visualization is developed, integrate it into the portal.
- Incorporate user feedback on the cost optimization tool and provide lessons learned and training.

Science Use Case Plans

1. **A Cloud-Based Framework for Visualization and Analysis of Big Geo Data** (Varun Chandola, University at Buffalo). Continue development of an iGlobe-based application to support sustainability research at UB (http://www.buffalo.edu/ubnow/stories/2019/08/renschler-great-lakes.html). The plan is to create a data and computing infrastructure, powered by Aristotle, to support a large collection of researchers, community stakeholders, and sustainability and resiliency planners, working towards creating partnerships for the sustainable development of the Lower Great Lakes area. During the inaugural NSF planning workshop, held in August 2019, we demonstrated the first version of the Aristotle platform. The next steps are to refine the platform and allow researchers to contribute data and run analyses.
2. **Global Market Efficiency Impact** (Dominik Roesch, University at Buffalo). Work with a UB student to automate part of the financial data framework hosted on Aristotle and integrate it with CCR OnDemand, OSC's Open OnDemand interactive HPC via the web. Continue to mentor 4 PhD student researchers. Resubmit research on the impact of arbitrage using Thompson Reuters Tick History data to the Journal of Financial Economics and write 3 new papers on financial market frictions and learning from stock price; global market efficiency; and, the impact of jumps in prices on holding period returns and volatility.
3. **Application of the Weather Research and Forecasting (WRF) Model for Climate-Relevant Simulations in the Cloud** (Sara C. Pryor, Cornell University). Continue to run high-resolution numerical simulations to improve the understanding of wind climate variability and change with a specific focus on applications to the wind energy industry. Continue to conduct high-resolution numerical simulations to improve the understanding of local climate perturbations resulting from the action of wind turbines in harnessing the kinetic energy of the

atmosphere and converting it into (carbon-free) electricity.

4. **Transient Detection in Radio Astronomy Search Data** (James Cordes, Cornell University). Build and deploy a Docker container with all the code, including the new Python 3-enabled version of PRESTO, for migrating to public cloud; test and run on at least two public clouds. Add a flood-fill algorithm. Build Machine Learning approaches to Fast Radio Burst detection and test on archived PALFA data.
5. **Water Resource Management Using Python Rhodium Framework** (Patrick Reed, Cornell University). Continue development of the Aristotle MPI cluster, i.e., MPI in a container. Benchmark the scalability of the water resource management software stack and investigate whether many containers can be spun up across multiple clouds, including bursting to AWS. Achieve efficient multi-node support and benchmark the Parallel Platypus VM for scaling. Build and test two additional software batches.
6. **Quantifying the Metabolic Determinants of the Fate of Probiotic Bacteria Administered to an Animal Gut Microbiota** (Angela Douglas, Cornell University). Identify microbial metabolic factors that influence the capacity of single “probiotic” bacteria to colonize an established microbiota of 1 to 5 members in the Drosophila gut, using SteadyCom on our previously constructed multi-compartment metabolic models. Determine the metabolic traits of individual microorganisms and among-microbial interactions that exclude probiotic bacteria. Integrate dynamic FBA into the SteadyCom colonization simulations to investigate changes to animal gut nutrient composition in response to among-microbial metabolite exchange with and without the probiotic strain. Use high-nutrient and low-nutrient conditions to determine impact of gut nutrient status on colonization or exclusion of probiotic bacteria. Design empirical colonization experiments informed by predictions from metabolic models to define the metabolic conditions that promote probiotic colonization and beneficial function.
7. **Multi-Sourced Data Analytics to Improve Food Production and Security** (Kate McCurdy, Sedgwick Reserve; Elizabeth Grafton-Cardwell, UC Lindcove Research & Extension Center; Chandra Krintz, UCSB SmartFarm with Cal Poly, San Louis Obispo and Fresno State). Develop sustainable land use practices at Sedgwick that employ livestock as part of the management lifecycle and install new monitoring infrastructure. Phase 2 infrastructure has been deployed at Lindcove in time for the 2019 citrus “frost season” and is in testing. During the next year, the science team is anticipating its first full season frost monitoring data sets. In addition, the science team is partnering with a USDA-funded effort to study HLB citrus greening remediation using screen houses for citrus products. Called CUPS (Citrus Under Protective Screening), the science team will be providing the instrumentation and analytics necessary to evaluate the first CUPS installation (at scale) in California. The installation is planned for September 2019 with instrumentation expected to be in place by late 2019 or early 2020. Aristotle will be providing the computational infrastructure to the team that is necessary to analyze the effects of CUPS on citrus production.

Plans to Disseminate Results

The Aristotle team will continue to engage the cyberinfrastructure community through presentations and dialogue at CASC, SC19 and PEARC20, and future scientific meetings such as the 4th ACM/IEEE Symposium on Edge Computing, November 7-9, 2019, Washington, DC where Aristotle co- PI Wolski and use case scientist Krintz will present “FaaSIoT: Portable, Multi-scale Functions-as-a-Service for IoT.” Aristotle PI/co-PIs will continue to respond to inquiries regarding the project and keep leadership at XSEDE and relevant NSF projects abreast of new developments in the federated cloud model. The Aristotle portal will highlight project results, and be updated regularly with news, events, and additions to the user guide and build-your-own-cloud federation documentation.

Products

Books

Angela E. Douglas (2018). *Fundamentals of Microbiome Science: How Microbes Shape Animal Biology* Princeton University Press. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = No ; ISBN: 9780691160344

Book Chapters

Sara C. Pryor & Andrea N. Hahmann (2019). Downscaling wind. *Oxford Research Encyclopedias: Climate Science* Hans Von Storch. Oxford University Press. . Status = PUBLISHED; Acknowledgement of Federal Support = Yes ; Peer Reviewed = No ; DOI: 10.1093/acrefore/9780190228620.013.730.

Inventions

Journals or Juried Conference Papers

Arun Sharma, Syed Mohammed Arshad Zaidi, Varun Chandola, Melissa R. Allen & Budhendra L. Bhaduri (2018). WebGlobe – A cloud-based geospatial analysis framework for interacting with climate data. *Juried Conference Paper*. 42. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; DOI: 10.1145/3282834.3282835

Chandra Krintz, Rich Wolski, Nevena Golubovic & Fatih Bakir (2018). Estimating outdoor temperature from CPU temperature for IoT applications in agriculture. *Juried Conference Paper*. . Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; DOI: 10.1145/3277593.3277607

Craig A. Stewart, David Y. Hancock, Julie Wernert, Thomas Furlani, David Lifka, Alan Sill, Nicholas Berente, Donald F. McMullen, Thomas Cheatham, Amy Apon, Ron Payne & Shawn D. Slavin (2019). Assessment of financial returns on investments in cyberinfrastructure facilities: A survey of current methods. *Juried Conference Paper*. . Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; DOI: 10.1145/3332186.3332228

Fatih Bakir, Rich Wolski, Chandra Krintz & Gowri Ramachandran (2019). Devices-as-services: Rethinking scalable service architectures for the internet of things. *Juried Conference Paper*. . Status = AWAITING_PUBLICATION; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes

Frederick Letson, Rebecca J. Barthelmie & Sara C. Pryor (2019). RADAR-derived precipitation climatology for wind turbine blade leading edge erosion. *Journal*. . Status = AWAITING_PUBLICATION; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; DOI: 10.5194/wes-2019-43

Gareth George, Rich Wolski, Chandra Krintz & John Brevik (2019). Analyzing AWS spot instance pricing. *Juried Conference Paper*. 222. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; DOI: 10.1109/IC2E.2019.00036

Jeanette M. Sperhac, Benjamin D. Plessinger, Jeffrey T. Palmer, Rudra Chakraborty, Gregary Dean, Martin Innus, Ryan Rathsam, Nikolay Simakov, Joseph P. White, Thomas R. Furlani, Steven M. Gallo, Robert L. DeLeon, Matthew D. Jones, Cynthia Cornelius & Abani Patra (2018). Federating XDMoD to monitor affiliated computing resources. *Juried Conference Paper*. . Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes

Kee H. Chung, Albert J. Lee & Dominik Rosch (). Tick size, liquidity for small and large orders, and price informativeness: Evidence from the Tick Size Pilot Program. *Journal*. . Status = AWAITING_PUBLICATION; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; DOI: 10.2139/ssrn.3220470

Kyle Carson, John Thomason, Rich Wolski, Chandra Krintz & Markus Mock (2019). Mandrake: Implementing durability for edge clouds. *Juried Conference Paper*. . Status = AWAITING_PUBLICATION; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes

Michael Zhang, Chandra Krintz, Rich Wolski & Markus Mock (2019). Seneca: Fast and low cost hyperparameter search for machine learning models. *Juried Conference Paper*. . Status = AWAITING_PUBLICATION; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes

Nana Y.D. Ankrah, Bassem Chouaia & Angela E. Douglas (2018). The cost of metabolic interactions in symbioses between insects and bacteria with reduced genomes. *Journal*. 1433. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; DOI: 10.1128/mBio.01433-18

Nevena Golubovic, Chandra Krintz, Rich Wolski, Balaji Sethuramasamyraja & Bo Liu (2019). A scalable system for executing and scoring K-means clustering techniques and its impact on applications in agriculture. *Journal*. 6 (3/4), 163. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; DOI: 10.1504/ijbdi.2019.10021277

Nevena Golubovic, Rich Wolski, Chandra Krintz & Markus Mock (2019). Improving the accuracy of outdoor temperature prediction by IoT devices. *Juried Conference Paper*. . Status = AWAITING_PUBLICATION; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes

Sara C. Pryor, Tristan J. Shepherd, Rebecca J. Barthelmie, Andrea N. Hahmann & Patrick Volker (2019). Wind farm wakes simulated using WRF. *Journal*. 1256 (1), 10.1088.1742-6596/12. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; DOI: doi

Syed Mohammed Arshad Zaidi, Varun Chandola, Melissa R. Allen, Jibonananda Sanyal, Robert N. Stewart, Budhendra L. Bhaduri & Ryan A. McManamay (2018). Machine learning for energy-water nexus: challenges and opportunities. *Journal*. 2 (3), 228. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; DOI: 10.1080/20964471.2018.1526057

Wei-tsung Lin, Fatih Bakir, Chandra Krintz & Markus Mock (2019). Data repair for distributed, event-based IoT applications. *Juried Conference Paper*. 139. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; DOI: 10.1145/3328905.3329511

Licenses

Other Conference Presentations / Papers

Chandra Krintz (2018). *Adventures and opportunities in cyber-physical systems and research*. 2018 International Conference on Computer Aided Design. San Diego, CA. Status = OTHER; Acknowledgement of Federal Support = Yes

Sayed Mohammed Arshad Zaidi, Varun Chandola, Melissa R. Allen & Bhudhendra L. Bhaduri (2018). *Anomaly detection in Energy-Water Nexus: webGlobe – A cloud-based geospatial analysis framework for interacting with climate data*. American Geophysical Union Fall 2018 Meeting. Washington, DC. Status = OTHER; Acknowledgement of Federal Support = Yes

Terrence Hendershott, Dmitry Livdan & Dominik Roesch (2018). *Asset pricing: A tale of night and day*. American Finance Association 2019 Annual Meeting, Jackson Hole Finance Conference & Desmarais Global Finance Research Centre. Vienna, Jackson Hole, WY & Montreal. Status = OTHER; Acknowledgement of Federal Support = Yes

Rich Wolski, Chandra Krintz & Wei-tsung Lin (2018). *CSPOT: A serverless platform of things*. University of California, Santa Barbara Computer Science Technical Report. Santa Barbara, CA. Status = OTHER; Acknowledgement of Federal Support = Yes

Matt Baughman, Simon Caton, Christian Haas, Ryan Chard, Rich Wolski, Ian T. Foster & Kyle Chard (2019). *Deconstructing the 2017 changes to AWS spot market pricing*. ScienceCloud@HPDC. Boulder, CO. Status = OTHER; Acknowledgement of Federal Support = Yes

Rich Wolski (2019). *Devices-as-services: Rethinking scalable service architectures for the Internet of Things*. 2019 USENIX Annual Technical Conference. Reston, VA. Status = OTHER; Acknowledgement of Federal Support = Yes

Susan Mehringer (2018). *Federations: Lessons learned*. Coalition for Academic Scientific Computation Fall Meeting. Alexandria, VA. Status = OTHER; Acknowledgement of Federal Support = Yes

Richard Knepper, Resa Reynolds, Eric Coulter & Steve Bird (2019). *Gathering requirements for a campus cloud toolkit*. Practice and Experience in Advanced Research Computing (PEARC19). Chicago, IL. Status = OTHER; Acknowledgement of Federal Support = Yes

Shami Chatterjee (2018). *New results in radio astronomy: Fast radio bursts and transients*. International Astronomical Union General Assembly. Vienna. Status = OTHER; Acknowledgement of Federal Support = Yes

Tristan Shepherd, Bennett Wineholt, Rebecca J. Barthelmie & Sara C. Pryor (2019). *Quantifying weather and*

climate simulation reproducibility in the cloud. 99th American Meteorological Society Annual Meeting: 5th Symposium on High Performance Computing for Weather, Water, and Climate. Phoenix, AZ. Status = OTHER; Acknowledgement of Federal Support = Yes

Richard Knepper (2019). *Red Cloud and Aristotle: Campus clouds and federation*. Humanware Advancing Research in the Cloud (HARC) Workshop. Chicago, IL. Status = OTHER; Acknowledgement of Federal Support = Yes

Chandra Krintz (2018). *SmartFarm: IoT systems that simplify and automate agriculture analytics*. 8th International Conference on Internet of Things (IoT 2018). Santa Barbara, CA. Status = OTHER; Acknowledgement of Federal Support = Yes

Shami Chatterjee (2018). *The dwarf galaxy host of a repeating fast radio burst*. International Astronomical Union General Assembly. Vienna. Status = OTHER; Acknowledgement of Federal Support = Yes

Rebecca J. Barthelmie, Sara C. Pryor & Tristan J. Shepherd (2018). *Wakes from wind turbine array*. National Renewable Energy Laboratory. Golden, CO. Status = OTHER; Acknowledgement of Federal Support = Yes

Rebecca J. Barthelmie, Sara C. Pryor & Tristan J. Shepherd (2018). *Wakes from wind turbine array*. National Renewable Energy Laboratory. Golden, CO. Status = OTHER; Acknowledgement of Federal Support = Yes

Sara C. Pryor, Tristan J. Shepherd, Melissa Bukovsky & Rebecca J. Barthelmie (2019). *Wind energy scenarios for climate change mitigation*. The Scenarios Forum 2019. Denver, CO. Status = OTHER; Acknowledgement of Federal Support = Yes

Other Products

Audio or Video Products.

1. Audio or Video Products

CENTAURUS: K-Means Cluster as a Service video (2019) demonstrates the balancing of a scalable clustering analysis workload between two Aristotle clouds. Download the video from https://federatedcloud.org/images/Cloud_demo_openstack.mov. An IEEE paper describes CENTAURUS: https://ieeexplore.ieee.org/document/8328527.

2. Audio or Video Products

Aristotle co-PI Rich Wolski interview by Washington DC's The Hill (2018) highlights the benefits of Smart Farm Technologies. See the video at https://thehill.com/hilltv/boundless/421077-smart-farming-will-help-cutdown-on-food-waste-says-computer-scientist.

3. Audio or Video Products

Aristotle WRF Container video (2018) demonstrates the pulling or caching of Docker images to get the WRF app running quickly. See the video at https://federatedcloud.org/images/Aristotle%20WRF%20Container%20Demo.mp4.

4. Audio or Video Products

Aristotle Use Case Scientists Angela Douglas video (2018) describes her research into how host-microbe interactions affect metabolism and nutrition. See the interview at https://www.youtube.com/watch?v=jup7FXzw7Tw.

Other Publications

Patents

Technologies or Techniques

The Aristotle project team and science use case researchers developed the following technologies and techniques:

Aristotle AWS SpotPrice Visualizer – tool to predict AWS spot instance prices (beta)(https://ieeexplore.ieee.org/document/8790118).
CSPOT – portable, multi-scale Functions-as-a-Service (FaaS) system for implementing IoT applications (https://www.cs.ucsb.edu/sites/cs.ucsb.edu/files/docs/reports/master_2.pdf).
Devices-as-Services – new "flipped" client-server model for IoT in which devices at the edge are servers that provide nanoservices, which applications in the cloud (the clients) compose for their implementations (https://www.usenix.org/system/files/hotedge19-paper-bakir.pdf).
Fast Radio Burst (FRB) Pipeline – new pipeline to detect single pulse candidates that may be FRB sources in radio astronomy data.
Federated Open XDMoD – tool to monitor affiliated (cloud) computing resources (https://federatedcloud.org/papers/hpcmaspa-ieee-approved-PID5492135.pdf).
K-means clustering analysis system – scalable system for executing and scoring K-mean cluster techniques; runs as a cloud service on Aristotle and Jetstream(https://sites.cs.ucsb.edu/~ckrintz/papers/centaurus-journal18.pdf).
Mandrake – software infrastructure for edge clouds (private clouds located at the network edge), designed to provide reliable, "lights out" unattended operation and application hosting in IoT deployments (https://sites.cs.ucsb.edu/~rich/publications/edge-19.pdf).
Multicloud run method - uses Python, Celery distribute task queue, and other tools to run applications across multiple cloud sites.
Temperature prediction methods - new methods for improving the accuracy of outdoor temperature prediction by IoT devices (https://sites.cs.ucsb.edu/~rich/publications/iciot-19.pdf).
Seneca – fast and low cost hyperparameter search method for machine learning models (https://sites.cs.ucsb.edu/~ckrintz/papers/ieee_cloud19.pdf).
Singularity and container images – new images that create reproducible research workflows which due to their portability can be shared broadly across institutions and disciplines, and run on any cloud (Aristotle, NSF clouds, or public clouds).
webGlobe – a cloud-based geospatial analysis framework for interacting with climate data(https://dl.acm.org/citation.cfm?id=3282835).
Wind turbine blade analysis framework – robust, flexible framework for generating an observationally constrained georeferenced assessment of precipitation-induced wind turbine blade erosion (https://www.wind-energ-sci-discuss.net/wes-2019-43/wes-2019-43.pdf).

Thesis/Dissertations

Websites
Aristotle Cloud Federation
https://federatedcloud.org

The Aristotle Cloud Federation portal was updated regularly to feature new web content and user guide documentation as well as Aristotle science use case accomplishments, plans, and products. Quarterly National Science Foundation project reports are also available on the portal; they provide detailed descriptions of the project's activities, challenges, and accomplishments.

Participants/Organizations

Research Experience for Undergraduates (REU) funding

Form of REU funding support:	REU supplement
How many REU applications were received during this reporting period?	30
How many REU applicants were selected and agreed to participate during this reporting	5

period?

REU Comments:

What individuals have worked on the project?

Name	Most Senior Project Role	Nearest Person Month Worked
Lifka, David	PD/PI	1
Furlani, Thomas	Co PD/PI	1
Wolski, Richard	Co PD/PI	3

Full details of individuals who have worked on the project:

David A Lifka
Email: lifka@cac.cornell.edu
Most Senior Project Role: PD/PI
Nearest Person Month Worked: 1

Contribution to the Project: Programmatic oversight of the Aristotle Cloud Federation project ensuring deliverables outlined in the program execution plan are met on schedule.

Funding Support: No funding support from other projects used for this award.

International Collaboration: No
International Travel: No

Thomas R Furlani
Email: thomas.furlani@roswellpark.org
Most Senior Project Role: Co PD/PI
Nearest Person Month Worked: 1

Contribution to the Project: Programmatic oversight of the UB subaward, and continuous interaction with the entire Aristotle technical team.

Funding Support: No funding support from other projects used for this award

International Collaboration: No
International Travel: No

Richard Wolski
Email: rich@cs.ucsb.edu
Most Senior Project Role: Co PD/PI
Nearest Person Month Worked: 3

Contribution to the Project: Architected the deployment at UCSB for production Aristotle services. Developed the AWS spot market prediction system. On-boarded new Sedgwick science team efforts and other science use cases.

Funding Support: No funding support from other projects used for this award.

International Collaboration: No
International Travel: No

What other organizations have been involved as partners?

Name	Type of Partner Organization	Location
Dartmouth	Academic Institution	Hanover, NH

Full details of organizations that have been involved as partners:

Dartmouth

Organization Type: Academic Institution
Organization Location: Hanover, NH

Partner's Contribution to the Project:
In-Kind Support
Collaborative Research

More Detail on Partner and Contribution: Dartmouth is collaborating with the Aristotle team to test our Federated Cloud resources and documentation

What other collaborators or contacts have been involved?
Nothing to report

Impacts

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

The Aristotle project is advancing the knowledge of federated cloud computing and its potential role as a campus bridging paradigm. By building and deploying a federated cloud model with the necessary allocations, accounting, and cloud metrics, Cornell University, University at Buffalo, and UC Santa Barbara are exploring how cloud resources can be effectively shared between campuses and their impact on researchers who require flexible workflows and analysis tools for large-scale data sets. The project serves as an important model for campus cyberinfrastructure that others may follow and provides elasticity by sharing resources, data, software, and tools that may not be available locally. The project is also exploring an allocation model that provides a fair exchange mechanism for resource access between institutions.

What is the impact on other disciplines?

Aristotle use case scientists are strategically exploring problems of increasing complexity and corresponding increases in data and, as a result, are advancing scientific knowledge. Data challenges from a diversity of disciplines (earth and atmospheric sciences, finance, chemistry, astronomy, civil engineering, genomics, and agriculture) are being addressed with collaborators from other academic institutions, public agencies, and research labs, as well as citizen scientists. The sharing of data infrastructure building blocks capacity and the movement of instances across institutional boundaries has the potential to create wider science collaborations and increased data sharing. The creation and performance testing of containerized applications can provide portability to from local cloud resources to other campus clouds, NSF clouds, or public clouds, thereby, producing a positive impact on research productivity in terms of agility and cost-effectiveness.

What is the impact on the development of human resources?

Aristotle is pioneering the concept of federated cloud computing for research which may ultimately increase the availability of on demand resources, data, and analysis tools that engage underrepresented groups. In addition, virtual laboratories in the cloud can enhance classroom learning. For example, Aristotle cloud resources were used solve large problems in a Computer Science class (CS293B) taught at UC Santa Barbara in winter 2019. The availability of campus-to-campus cloud computing may also spurn the development and dissemination of ready-to-launch VMs and containers with training software and tools preloaded. This could reduce the redundant development and preparation of educational material development and onsite computer labs administration, resulting in an increased focus on individual student learning needs. Fourteen REU students have participated in the Aristotle project to date.

What is the impact on physical resources that form infrastructure?

The federated cloud model may impact the physical resources that form infrastructure by reducing the number of computer labs required for learning. Campus clouds may be installed so that researchers can cost effectively use local cloud resources and, when more capacity is needed, move their application container to the most suitable campus, public, or NSF cloud resource. Ultimately, federated clouds will likely become complementary resources to high-end supercomputers, e.g., performing on demand iterative tasks, streaming IoT data, etc. Integrating federated clouds (campus, labs, etc.) into the national cyberinfrastructure ecosystem may increase the sharing of resources and, subsequently, reduce physical resource expenditures at the local level.

What is the impact on institutional resources that form infrastructure?

Aristotle will maximize institutional resources through federation with other institutions by (1) offloading variable computational and data analysis demands from local infrastructure, (2) starting coarsely parallel computations on demand, (3) providing heterogeneous instance types (CPUs, GPUs, etc.) and sizes to allow for unpredictable computational demand.

What is the impact on information resources that form infrastructure?

Aristotle's federated cloud model will facilitate (1) sharing of high-value processed datasets of general interest and separate data resources, (2) generation of reproducible pipelines in the form of VMs or VM configurations, and containers (3) access to multiple data sources, some of which are already in public and private clouds.

What is the impact on technology transfer?

Technologies generated by this project are open source, therefore technology transfer licensing, patent applications, etc. are not applicable. The project's collaborative relationships for R&D include Amazon Web Services, Dell, Globus, Google, Flexera/RightScale, Microsoft Azure, Red Hat, and Sylabs.

What is the impact on society beyond science and technology?

Aristotle use cases have the potential to impact wind turbine companies interested in harnessing the energy of the atmosphere and converting it into carbon-free electricity; policymakers regulating high-frequency trading; policymakers making water resources management decisions; manufacturers producing sustainable insect pest management products; farmers increasing yields and protecting the environment by accessing on demand soil, water, and crop sensor data to aid decision-making; and, orange growers trying to protect U.S. citrus trees from the Huanglongbing bacteria (citrus greening disease) that has devasted FL and is threatening CA.

Changes/Problems

Changes in approach and reason for change
Nothing to report.

Actual or Anticipated problems or delays and actions or plans to resolve them
Nothing to report.

Changes that have a significant impact on expenditures
Nothing to report.

Significant changes in use or care of human subjects
Nothing to report.

Significant changes in use or care of vertebrate animals
Nothing to report.

Significant changes in use or care of biohazards
Nothing to report.