![USDA United States Department of Agriculture — Agricultural Research Service]

SCINet Newsletter: October 2021

NEWS

Storage Update

There have been significant improvements to SCINet data storage and management capabilities including increased storage capacity and better data safety. To take full advantage of these improvements, ARS researchers need to update how they store and work with data on SCINet infrastructure. The diagram below outlines suggested best practices for managing data storage for compute jobs on ARS's HPC clusters. Juno, ARS's new storage system, should be used for all long-term data storage because data and files saved on Juno are routinely backed up. Data and files stored on Ceres and Atlas are not backed up. For more information read the new standard operating procedure for basic data management on SCINet.
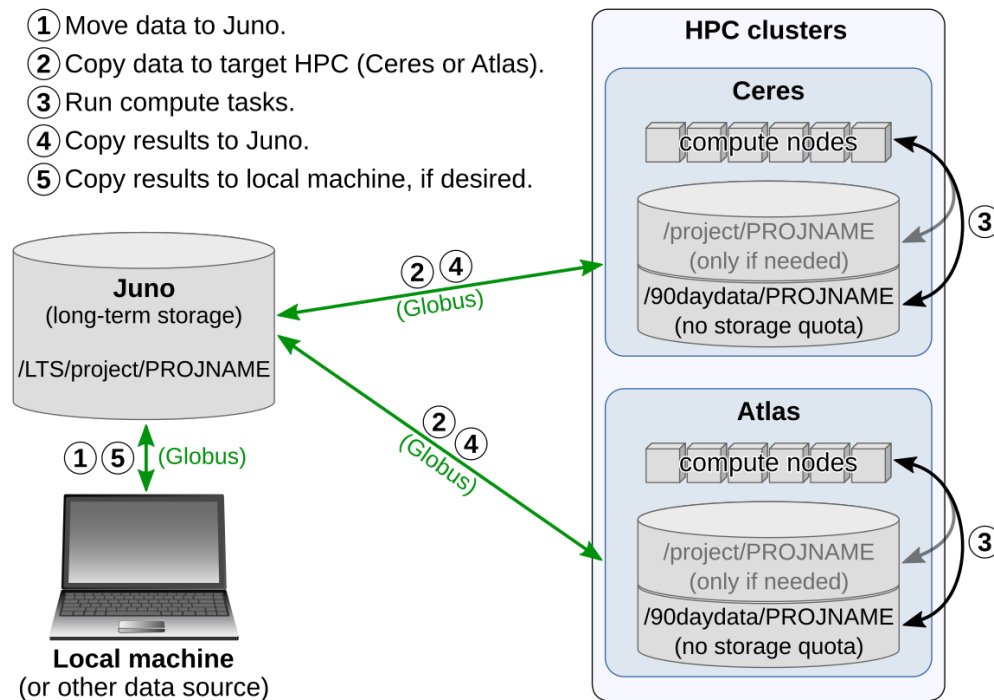
SCINet Policy Updates

The SCINet Policy Committee has been busy helping make a safer SCINet and encourage a positive SCINet experience. The committee is made up of scientists, bioinformaticians and IT professionals across ARS. If you have policy questions or are interested in joining the committee, please contact the Policy Committee Chair. Please take a moment to read the newest policy updates.

SCINet Fellows Conference

SCINet and AI Center of Excellence Fellows will be presenting their innovative research utilizing SCINet infrastructure on November 9-10, 2021. Fellow will deliver short talks followed by panel discussions involving National Program Leaders and Research Leaders across a variety of topics. Keep an eye on the upcoming events page for more information.

3D Proteins in Agriculture Research Community

Did you know that at least 5% of the ARS research community is interested in protein folding technologies? A recent survey was sent out to ARS scientists to measure interest in protein folding capabilities and training. Over 100 respondents were interested in training and support related to protein folding. If you didn't get a chance to respond to the survey, but you are interested in learning more about protein folding technologies, please reach out to scinet-training@usda.gov to be part of the "unfolding" community.

Pollinator Workgroup Announcement

The SCINet Pollinator working group was initiated in February 2021 with a goal of promoting coordinated, computational research on managed and native pollinators. The group hosted a workshop in August to explore leveraging advances in data science to manage and conserve pollinators. To learn more about the working group and explore the exciting workshop content, visit the new Pollinator working group webpage at https://scinet.usda.gov/working-groups/pollinator.

SAC Seeking Representative

The Scientific Advisory Committee (SAC) supports gathering and sharing information to support research in the scientific computing and artificial intelligence communities. SAC members represent a breadth of scientific research at ARS. Membership includes two researchers from each of the five ARS geographic areas, a statistician member, and an "at-large" member that can be located in any of the ARS' geographic areas. The SAC divides

its work over various subcommittees: communications, education, planning, and membership.

Please send nominations for the "at-large" representative to [Kathy Yeater, SAC Membership Chair](). Self-nominations are encouraged. Please include a short paragraph of your background and interest in serving on the SAC. The current term for this open role expires in Spring 2022, with the opportunity to continue for a three-year term (Spring 2022 – Spring 2025). Please consider sharing your time and talents on the SAC.

RESEARCH SPOTLIGHT

Putting flowers on the map: Quantifying spatio-temporal patterns in floral resources for pollinators

By: Dr. Melanie Kammerer

Melanie is an ORISE Fellow within the SCINet Office. Melanie completed her PhD in Ecology at Penn State University where she studied landscape and climate drivers of Wild Bee Communities. She leads the SCINet Pollinator Working Group.

Flowers provide critical nectar and pollen resources to support foraging, learning and memory, and reproduction of pollinating insects, but there are very few large-scale maps of floral resources for

pollinators. As part of my post-doctoral research supported by SCINet, I am developing a data-driven method to map floral resources at large geographic extents by scaling-up field observations of plant communities. My colleagues and I are integrating previously disparate datasets that describe 'who, when, where, and how much' flowering occurs in 23 habitat types in the Northeast USA.

Our approach is unique because it generates relatively fine-scale temporal predictions of floral area over large spatial extents, relevant for apicultural, agricultural, and conservation decisions. With this large, integrated dataset on flowering, for a given location in the Northeast, we can estimate monthly (or even weekly) quantities of flowers available for pollinators across a. Some wild-bee species are active for a very short segment of the growing season but forage up to 1-5km from their nests (or more in times of resource scarcity), necessitating a combination of fine temporal and broad spatial data. This work will enable researchers, land owners, and conservation professionals to identify temporal and spatial gaps in available flowers and design pollinator plantings to mitigate resource scarcity (recently highlighted as federal research priority in the 2021 USDA Annual Strategic Pollinator Priorities and Goals Report).

SCINet note: Our team used SCINet's Atlas HPC to create national habitat maps that we could combine with field surveys to map floral area. This workflow involved merging two pre-existing habitat maps (USDA NASS Cropland Data Layer with the LANDFIRE National Vegetation Classification). Using the Atlas 'big memory' nodes, we completed the geospatial merge for the conterminous United States in less than one day which will allow us to create annual habitat maps for multiple years (2010-2020 expected output).

Do you use SCINet for your research? Contact SCINet-Newsletter@usda.gov for a chance to be featured in the newsletter!

TRAINING

Training Opportunities

The Carpentries Workshops: Multi-day remote workshops covering Unix, git, and either R or Python from The Carpentries will be offered this Winter. Details will be posted on the Upcoming Events page. There is also an opportunity to become a Carpentries workshop instructor. This training teaches participants The Carpentries pedagogy and results in the ability to lead Carpentries workshops in Unix, git, R, Python, and more. To get on the waitlist for any of these courses, please email SCINet-training@usda.gov.

Coursera.org certified courses update: The SCINet Office and the AI Center of Excellence are excited to provide training opportunities through Coursera. Coursera licenses are available to ARS scientists and support staff to complete training focused on scientific computing and artificial intelligence. Successful completion of courses and

specializations will result in widely recognized certificates and credentials. Please visit the SCINet Coursera Training page to request a license by quarter.

Free Online Computational Training (Self-paced): Make use of your work-from-home time with computational training! A large list of free tutorials and courses has been compiled on the Free Online Training page. Training topic areas include Python, R, SAS, and MATLAB programming; statistics; data science concepts; AI and machine learning; GIS; Google Earth Engine; Git and GitHub; reproducibility, productivity, and integration management tools; and bioinformatics and ecology domain learning.

SCINet Online Science Tutorials: Browse our growing set of SCINet science tutorials created by ARS scientists and the SCINet Virtual Research Support Core. Our Science Tutorials page now includes Ceres Onboarding and Intro to Unix for new HPC users, two geospatial computing tutorials, a QTL Analysis tutorial for breeding, and machine learning training material.

Training opportunities are continuously being updated on the SCINet Upcoming Events webpage. For more information on any of the above trainings, registration questions or suggestions, please email SCINet-training@usda.gov.

SUPPORT

Getting Started with SCINet is as Easy as 1,2,3

1. Request a SCINet account to get started.

2. Read the SCINet FAQs covering general info, accounts/login, software, storage, data transfer, support/policy/O&M, parallel computing, and technical issues.

3. Register for a SCINet Forum account to connect to other users, ask questions, and learn how SCINet can enable your research.

P.S. Don't forget to change your password when logging in for the first time.

For technical assistance with your SCINet account, please email scinet_vrsc@usda.gov.

SCINet User Tips

By: Dr. Andrew Severin

Andrew manages the Genome Informatics Facility at Iowa State University. He leads a team of experts tasked with enabling USDA scientists to translate big data into informative data on their specific scientific questions. He is an interdisciplinary scientist working at the interface of genetics and bioinformatics. His academic background is in biochemistry with a Ph.D. in NMR spectroscopy.

Just today I had a problem where I was creating a one-liner on the unix command line and wanted to sort a table but keep the header info at the top. Turns out someone else had this problem. The solution involves creating a bash function that you can put in your.bashrc file. This function will now save the first line of the file you are sorting and print it at the top!

```
function body() {
# print the header (the first line of input)
# and then run the specified command on the body (the rest of the input)
# use it in a pipeline, e.g. ps | body grep somepattern
IFS= read -r header
printf '%s\n' "$header"
"$@"
}
```

Ever wanted to submit a job to the SLURM schedule and have it be dependent on the completion of a different job? Check out how here: Submitting dependency jobs in slurm (https://bioinformaticsworkbook.org/Appendix/HPC/SLURM/submitting-dependency-jobs-using-slurm.html#gsc.tab=0)

Waiting to learn Unix later but still want to use SCINet resources? Try out Galaxy: Getting started in Galaxy (https://github.com/Sivanandan/galaxy_demo)

Installation can be challenging. Conda environments can make it easier to install open source programs you love. Setting up conda on Ceres vs Atlas (https://github.com/j23414/conda_demo)

Do you have tips to share? Email them to SCINet-Newsletter@usda.gov to be included in future newsletters.

SCINet Corner: Third Thursdays

SCINet Corner is a VRSC moderated virtual space for people to share knowledge, discuss best practices, learn about new opportunities, and explore resources to support progress on their projects.

This reoccurring meeting occurs on the third Thursday each month. The next event is on Thursday, October 21st (10am PDT, 11am MDT, 12pm CDT, 1pm EDT). Meeting times may change. It is recommended to join the event via Google Chrome or Firefox.

Register at https://forms.gle/7DcBoBvbGcjQDBP38

Have a question that just can't wait? Want to see what other users are doing? Reach out to the ever-expanding SCINet Forum community for ideas, support, or just someone to bounce ideas off of at https://forum.scinet.usda.gov/**.**

CONNECT

The SCINet Team

Every newsletter highlights SCINet community members as a way to connect the ARS scientific computing community. This issue highlights the dynamic SCINet Office team and some of ARS's newest ORISE Fellows contributing to the SCINet mission. To see all the SCINet community and review past newsletters, visit the Newsletter Archive.
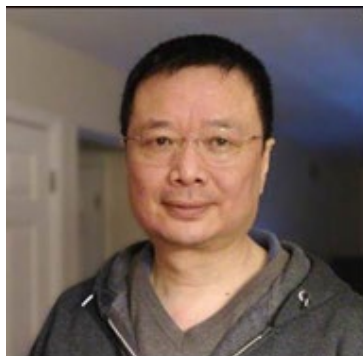
Haitao Huang, Computational Biologist (Data Scientist), SCINet Office- Haitao has extensive experience developing and modifying computer code for scientific projects. Most recently, he developed code for simulation models of herbaceous and woody species at transition zones, and for soil water dynamics and seedling establishment in grassland. He recently joined two SCINet HPC projects: the Geospatial Common Data Library project for frequently accessed data from USDA, USGS, NASA and Google Earth Engine, and the VSV Grand Challenge project to predict the spread of the Vesicular Stomatitis Virus. Haitao holds a Bachelor's degree in Engineering and a Master's degree in Computer Science.

Dr. Heather Savoy, Computational Biologist (Data Scientist), SCINet Office- Heather's research interests include applying informatics methods to multidisciplinary agro-ecosystem problems and building data science software tools for geospatial research. She is currently contributing to the Predictive Disease Ecology Grand Challenge project and the SCINet Geospatial Working Group's Geospatial Common Data Library project. She received her Ph.D. in Civil and Environmental Engineering with an emphasis in Computational Data Science and Engineering from the University of California Berkeley in 2017. She also holds a B.S. in Environmental Science with a minor in Computational

Mathematics from the Florida Institute of Technology. She was previously an ARS Biologist (Computational Bioinformatics) in Las Cruces, NM.

Dr. Brian Stucky, AI Computational Biologist, (Data Scientist) SCINet Office- Brian has a life-long fascination with computer science, biology, and agriculture, and his professional interests lie at the intersection of those fields. In particular, his work focuses on using tools and techniques from artificial intelligence, data science, statistics, and software engineering to help advance biological and agricultural research, and he enjoys helping other scientists apply new computational methods to their research. Brian has a Bachelor's degree in Computer Science and a Ph.D. in Ecology and Evolutionary Biology from the University of Colorado.

Dr. Heather Jones, Training Coordinator, SCINet Office- Heather supports SCINet training activities and is the SCINet newsletter editor. She earned her B.S. in Biology and Ed.D. in Educational Leadership from the University of South Florida. Heather's professional interests include empowering professionals through training and collaboration in virtual environments. She has a background in science curriculum, training program development, and project management across public, private, and government industries.

Fellow Focus

ORISE Fellows bring a fresh perspective to the ARS scientist community, while allowing these Fellows to learn through hands-on research under our top scientists. SCINet funds a number of ORISE Fellows each year with the mission to support research using high performance computing and computational science. If you are interested in learning more about their skillset, please reach out to their mentors.

Dr. Michael A. Alcorn, ORISE Fellow, SCINet Office-
Michael recently graduated from Auburn University with a
Ph.D. in Computer Science and Software Engineering and a
Graduate Minor in Mathematics. His research focused on
deep learning and machine learning, with projects ranging
from multi-agent spatiotemporal modeling, distribution
estimation, and evaluating the robustness of computer vision
models with 3D graphics. In addition, Michael interned with
Adobe Research and the Cleveland Indians where he
developed deep learning models for a variety of applications.
Prior to starting his Ph.D., Michael worked as a Machine
Learning Engineer/Data Scientist at Red Hat applying natural
language processing and econometrics techniques. Working under the mentorship of Brian
Stucky and Deb Peters, Michael is looking forward to applying his machine learning skills in
the agriculture domain and supporting scientists in integrating machine learning into their
own research.

**Dr. Andrew Oliver, ORISE Fellow, Western Human
Nutrition Research Center-** Andrew obtained his Ph.D. in
2021 from the University of California, Irvine, where he
studied how diet and life style shape the human microbiome.
He now works together with Dr. Danielle Lemay exploring
machine learning as a tool to better understand how diet
impacts human health. Andrew's research leverages data
generated from the USDA's Nutritional Phenotyping study
(FL100), which measured a variety of outcomes related to immune, metabolic, and
microbiome functioning from 393 study participants. In addition to research, Andrew plans
to develop instructional materials to make machine learning more accessible to nutrition
scientists.

**Do you know someone that might be interested in becoming an ORISE Fellow with
SCINet or the AI Center of Excellence? All opportunities have been recently updated
with an FY22 deadline. For more information on the positions including deadlines
and potential start dates, visit: https://www.zintellect.com/Catalog and enter the
keyword: SCINet.**

Contribute

Do you use SCINet for your research? We would love to share your story! Email SCINet-
Newsletter@usda.gov to contribute content, ask questions, or provide feedback on the
SCINet newsletter or website.

SCINet Leadership Team

Deb Peters, Acting Chief Science Information Officer

Rob Butler, Acting SCINet Project Manager

Adam Rivers, Science Advisory Committee (SAC) Chair

Steve Kappes, Associate Administrator