![USDA United States Department of Agriculture Agricultural Research Service](logo)

SCINet Newsletter: April 2022

Research Spotlight | **News** | **Training** | **Support** | **Connect**

RESEARCH SPOTLIGHT

Using SCINet HPC Resources to Develop Efficient UAS Imagery Analysis Workflows for Rangeland Management
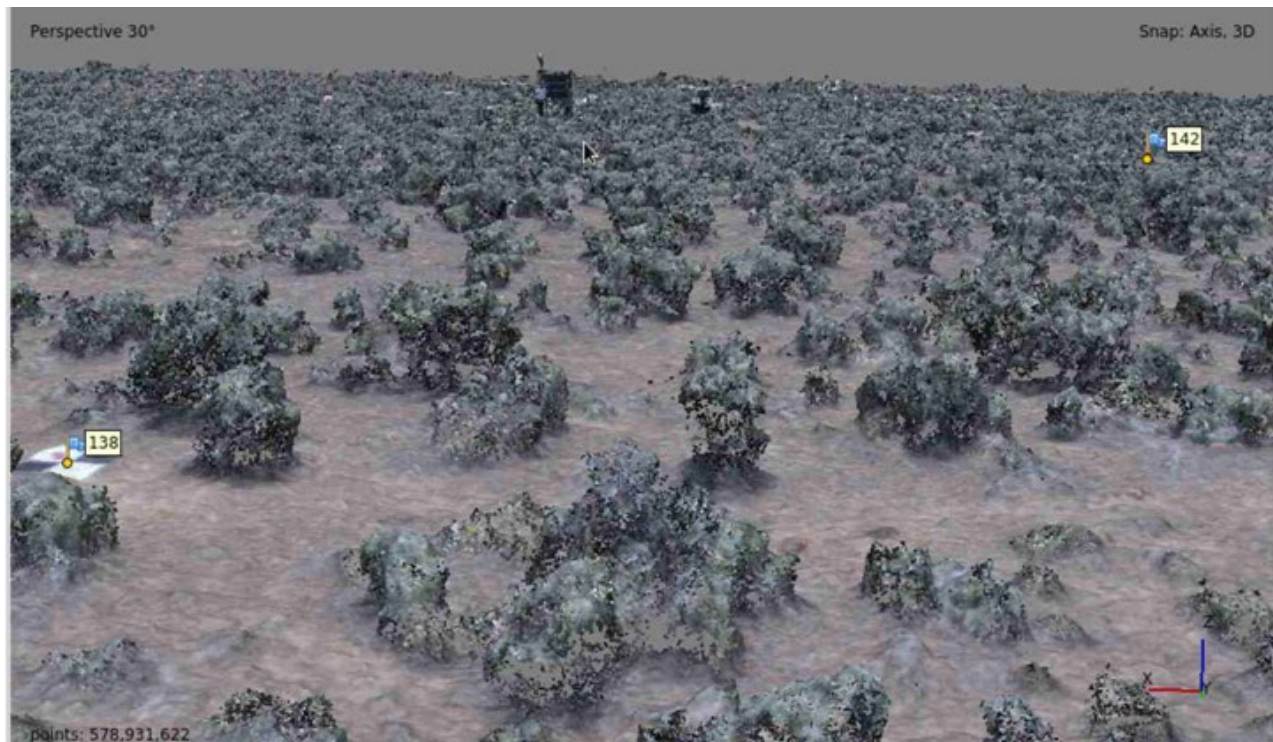
Image 1. The image illustrates the use of the Atlas Open OnDemand web interface to conduct the interactive workflow step of inspecting a dense point cloud created using Structure from Motion (SfM) analysis of UAS imagery. This dense point cloud represents a Wyoming big sagebrush plant community located at the Reynolds Creek Experimental Watershed in southwestern Idaho (Great Basin LTAR site). Shrubs illustrated here are about 0.25 to 0.5 m tall. Note the ground control point (GCP) target in the foreground and eddy covariance/phenocam tower in the background.

Dr. Pat Clark, Rangeland Scientist, USDA-ARS Northwest Watershed Research Center, Boise, Idaho.

Dr. Clark is developing and evaluating UAS imagery collection and analysis workflows with an eye towards augmenting or replacing field biomass measurements on spatially-extensive rangelands.

Working in several shrub-steppe vegetation types of the Great Basin region, Dr. Clark is using imagery collected using unmanned aerial systems (UAS) to develop digital surface, digital terrain, and canopy height models to intensively measure vegetation height, volume, and density. Plant seasonal production, fuel load, and biomass are then estimated from these vegetation structural measurements. Python scripts, originally developed by Andrew Cunliffe and colleagues at Exeter University (UK), are being used by Dr. Clark to automate imagery processing using software on the Atlas HPC. While a large portion of this analysis workflow can be accomplished with batch scripts, some steps require dynamic interaction with the analyst. Dr. Clark is using the Open OnDemand system available on Atlas to accomplish these interactive tasks.

Preliminary results are promising, but a considerable amount of evaluation and validation remain to be conducted. Expected products from this research include automated workflows, coded as python scripts, for conducting Structure from Motion (SfM), orthorectification, and mosaicking analyses. Additional workflow and model products will then enable a user to estimate production, fuel load, and biomass for rangeland plant functional types based on SfM-derived vegetation structure. This research holds promise for providing natural resource managers, fire and fuels specialists, and livestock producers with an efficient method to estimate rangeland biomass parameters using affordable, off-shelf UAS platforms and validated processing and analysis workflows.

NEWS

Guidemaker: Software for CRISPR Based Gene Function Discovery

CRISPR-Cas is a powerful gene-editing technology that works like miniature molecular scissors – it can precisely "snip" a piece of DNA, such as a target gene, out of an organism's genome. CRISPR-Cas gene removal, or "knockout", can help us learn about the function of a gene. For example, suppose we want to know if a particular gene confers antibiotic resistance in a strain of *Escherichia coli*. If we remove that gene from the bacterial genome and find that the bacteria are no longer resistant to the antibiotic, we have strong evidence that the target gene was responsible. How, though, does CRISPR-Cas know which gene to remove? The answer is that we must provide a custom-designed "guide RNA" molecule that allows CRISPR-Cas to find the target gene. Designing these guide RNAs can be quite difficult, however.

Dr. Adam Rivers, in collaboration with researchers at the University of Florida, has created user-friendly software called "GuideMaker" that can rapidly create guide RNAs for every gene in an organism's genome. Returning to our *E. coli* example, we could use GuideMaker to design guide RNAs that allow us to knock out *every* gene in the bacteria's genome simultaneously, so that we end up with a population of *E. coli* in which each bacterium has one gene removed. Rather than studying

the function of a single gene at a time, we can study the function of *all* genes in a single experiment. The software also works for eukaryotic cells.

GuideMaker is available online at https://guidemaker.app.scinet.usda.gov or in the Cyverse Discovery Environment. A command-line version of the software is available at https://guidemaker.org. GuideMaker can help bring the power of CRISPR-Cas to more scientists at ARS, and Dr. Rivers is happy to help ARS scientists get started with the method. A more detailed article about GuideMaker is available on the SCINet website and a peer-reviewed paper is in press at the journal GigaScience.

Arthropod Working Group Update

The ARS Arthropod Genomics Research (AGR) working group is hosting an international symposium, AGSx Virtual Symposium Spring 2022, with four sessions held once a month from February through May 2022. This symposium is facilitated by AGR lead Brad Coates, with Glenn Hanes providing technical expertise. AGR members Brenda Oppert and Lindsey Perkin and external collaborators organized and led respective scientific sessions: "Insect Genomic Technologies to Improve Food Applications", the "Honeybee Workshop", "Arthropod Genomics and Genome Engineering", and "Application of New Genomic Tools and Techniques in Arthropods". There were 251 registered participants from 21 countries through session II, and video recordings for each session are being posted to the i5K YouTube channel. This is the 2nd annual AGR-sponsored event, following the inaugural AGSx Virtual Symposium Spring 2021. The online format provides opportunities to exchange ideas and tools among a diverse group of international scientists involved in genomics research, highlights outcomes of the ARS Ag100Pest Initiative, and promotes diversity and inclusion.

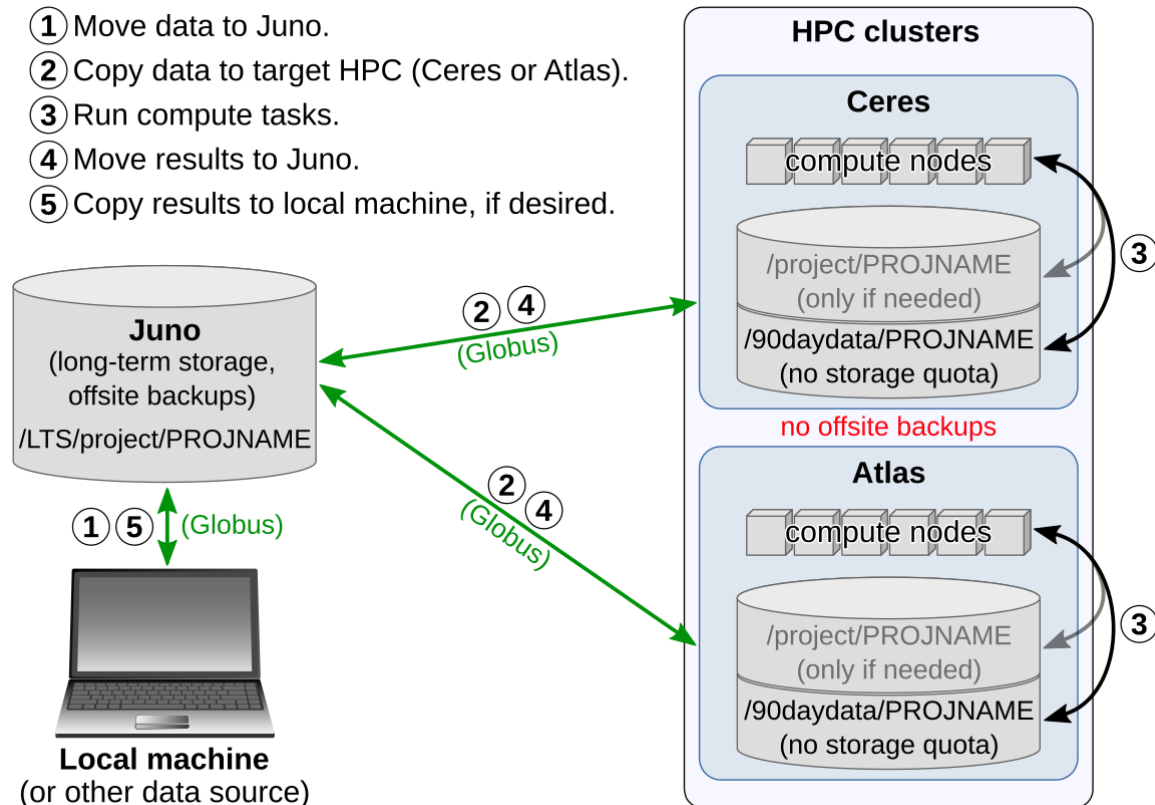# Transferring Data to and from the HPCs

Figure 1. Suggested workflow for using SCINet storage.

Thanks to recent SCINet storage improvements and streamlined storage policies and procedures, SCINet users now have more flexibility than ever before for working with and managing data and code on SCINet resources. Here, we briefly summarize SCINet storage options, suggested storage workflows, how to request more storage if you need it, and where you can go to learn more.

Storage options. SCINet storage comprises three storage locations. If you are already a SCINet user, you have likely interacted with one or both of our HPC (high-performance computing) clusters, called Ceres and Atlas. These clusters are our first two storage locations. Each cluster has a large (multiple petabytes), co-located storage array to support your computational work. Here are the key features of SCINet's HPC cluster storage:

- Storage on each cluster is split into two "bins": **/project** and **/90daydata**.

- Files in **/project** are retained indefinitely, but usage is governed by quotas that limit how much data a user can store in **/project**.

- Storage in **/90daydata** is not limited by quotas, but files are deleted after 90 days of inactivity.

- Files *are not* backed up externally. In the unlikely event of catastrophic system failure, data on the HPC clusters would be unrecoverable.

SCINet's other main storage system, called Juno, is a dedicated, high-capacity storage device that, unlike our HPC clusters, is not intended for computational work. Rather, it is a storage-only system. These are the key features of Juno storage:

- All files are retained indefinitely, but usage is governed by quotas that limit how much data a user can store on Juno.

- Files *are* backed up externally. In the unlikely event of catastrophic system failure, data backed up on Juno would be recoverable.

Suggested workflows. Figure 1 illustrates a suggested workflow for using SCINet storage. Data, code, and results that you *can't afford to lose* should be stored on Juno. On the HPC clusters, **/project** should *not* be used for long-term storage. Rather, files in **/project** should be limited to those required for active computational work on the cluster(s). Intermediate and temporary files, transient data, temporary computing environments, and other short-term project artifacts should be stored in **/90daydata**. If you implement this workflow, you will occasionally need to migrate data between cluster storage and Juno, and we recommend using Globus for these file transfers. These suggestions will work well for many SCINet projects, but we also recognize that one size does not fit all when it comes to scientific computing. We encourage you to contact the VRSC if you need help finding the best workflow for your project.

Getting more storage. When you set up a new project on SCINet, your project will be given 1 TB of storage in **/project** on each HPC cluster and on Juno. Although this storage amount will be adequate for many projects, some projects will require more storage space on one or more of our storage locations. We have recently implemented a completely new, streamlined process for requesting storage quota increases that only requires submitting a short form.

Where to learn more. To learn more about the topics discussed in this article, a good starting point is our recent SCINet Corner presentation that provides a complete overview of SCINet storage options and a live, step-by-step demonstration of our suggested SCINet storage workflow, including tips for efficient data transfer. The SCINet website provides more in-depth information about our suggested storage workflow, SCINet storage locations, and options for transferring data on SCINet. The VRSC can be contacted for specific questions.

SCINet and AI-COE Fellowship Update

We recently held a competition for ARS scientists wanting to mentor a postdoctoral fellow in the SCINet and AI-COE Fellows Program. Of the 64 proposals submitted, 8 were selected for funding by SCINet, and 11 for funding by the AI-COE. These fellows will join our current group to result in nearly 30 fellows in our fellowship program.

The goal of our program is to develop the next generation of ARS scientists with expertise in conducting and leading individual and collaborative research using computationally-intensive approaches. Each fellow is expected to participate in four types of activities: (1) individual computationally-intensive research with a mentor, (2) collaborative research with multiple ARS units, (3) co-leadership of a SCINet or AI COE working group with contributions to one of our workbooks, and (4) participation in the Fellows Program. In addition, each fellow will have the opportunity to take advantage of training courses that build computational literacy, such as in data science, AI, bioinformatics, and geospatial analyses, using our high performance computing clusters (Ceres, Atlas).

We congratulate the 19 ARS scientists who submitted successful proposals, and we look forward to the exciting and impactful research to be conducted by the fellows, their mentors, and the collaborative working groups. We expect to have another funding opportunity in FY23 so watch for that announcement in late 2022.

Do you know someone who might be interested in becoming an ORISE Fellow with a SCINet or the AI Center of Excellence mentor? There are a number of new opportunities being posted. Visit https://scinet.usda.gov/opportunities/postdocs for more information on the positions.

TRAINING

Training Report

Carpentries Winter/Spring Training Series: SCINet sponsored a series of Software and Data Carpentries trainings to support ARS researchers in developing scientific computing skills. The series included developing skills in UNIX, GIT, Python, R, OpenRefine, and SQL. Also, a new Data Carpentry Geospatial course is being added to the course offerings.

If you missed these events, then you can get on the waiting list to be notified when the courses are offered again this fall. Also, feel free to let us know what other training opportunities you are interested.

SCINet Corner: SCINet Corner is a moderated virtual spatial chat for people to share knowledge, discuss projects, and identify the resources needed to progress on their projects. Participants are encouraged to share information on tutorials, conferences, and other upcoming SCINet training opportunities.

Last month, Brian Stucky provided a short talk about long term storage, Juno and data transfer guidelines. View this recording. Also, sign up for the next SCINet Corner and come talk science with us!

Geospatial Training Series: The SCINet office sponsored a series of Geospatial Trainings to ARS staff. The trainings provided by Mississippi State University covered a range of topics including Developing a UAS program, UAS Data Wrangling, and Introduction to Image Processing. If you were

unable to attend these trainings, and would like to be contacted when offered again, please add your name to the waitlist.

Training Opportunities

Getting Started: With the expansive list of free training available online, finding the right training to meet your learning needs can be daunting. Take the first steps in getting started with the SCINet Introductory Learning Pathway. Learn about SCINet, how to sign up for an account, and what is possible when supported by SCINet infrastructure. Then dive in with hands-on tutorials available across multiple searchable platforms to find the information you need for just in time learning.

Certified Carpentries Instructors: If you are interested in joining the ARS Carpentries instructional team, this training will teach you Carpentries pedagogy and result in the ability to lead Carpentries workshops in Unix, git, R, Python, and more. If you are interested in becoming a Carpentries instructor, please reach out to scinet-training@usda.gov to get on the waitlist.

Courses by Mississippi State University: Mississippi State regularly offers Introduction to Atlas courses. Additionally, there are waiting lists available for several other courses, including an Intensive R course to help scientists with no R experience become familiar with the programming language and start performing statistical analyses in 4 days. Sign up to get notified when these courses are offered.

Coursera.org Courses Update: The SCINet Office and the AI-COE are excited to provide training opportunities through Coursera. These licenses are available to ARS scientists and support staff for training focused on scientific computing and artificial intelligence. Successful completion of courses and specializations result in widely recognized certificates and credentials from universities. Please visit the SCINet Coursera Training Page to request a license by April 15th for the quarter starting May 1st, 2022. Licenses turn over quarterly, although active learners can keep a license for longer than one quarter by request.

Training opportunities are continuously being updated on the SCINet Upcoming Training 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 complete your annual security training! This is required to maintain your account.

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

SCINet User Tip

Do you need to copy just a single file from a GitHub repository?

You don't need to download the entire repository to access a specific file locally. Here is how to do it.

1. Visit the chosen repository in your browser and navigate to the file of interest.
2. Click the "Raw" button in the upper right corner of the file display (see Figure).
3. Copy the URL.
4. From the terminal, run "wget URL_STRING", where URL_STRING is the URL you copied in step 3.

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

SCINet Corner: Third Thursdays Each Month

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 first Thursday each month. The next event is on Thursday, April 7th (1pm EST). It is recommended to join the event via Google Chrome or Firefox. In addition to meeting spaces to discuss research questions, there will be a meeting space to learn how to become a SCINet Carpentries instructor.

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 newest members of the SCINet Office. To see all the SCINet community and review past newsletters, visit the Newsletter Archive.

Dr. Ryan Lucas, Data Science Coordinator, SCINet Office- Ryan comes to the USDA ARS from the NOAA National Weather Service where he served as a Hydrologist with River Forecast Centers (RFC) in Portland, OR and Kansas City, MO. Ryan led Water Supply and Verification efforts at the Northwest RFC. Ryan earned his PhD and MSc in Environmental Systems from the University of California, Merced. Ryan also has a BS in Hydrologic Science from the University of California, Santa Barbara. Ryan has enjoyed working in both research and operational environments with work in the field, lab, and developing code. For the SCINet Office, Ryan will coordinate all of the scientific activities, and will interface with ARS scientists to address their needs for computational resources.

Moe Richert, Full Stack Developer, SCINet Office- Moe Richert has a background in ecology, communication, web development, and spatial analytics with a B.S. in Ecology from Auburn University and a M.S. in GIS and Web Map Programming from University of Wisconsin-Madison. Previously, he used these skills to run analytics, coordinate training and education activities, manage databases, and develop web applications to assist both government processes and public access to information. He is looking forward to working with the SCINet team on the front and backend of the SCINet web site and will be instrumental in developing a new website for the AI-COE.

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

Note: This newsletter is edited to comply with ARS editorial standards.

[SCINet Website](#)