

Introduction to the Research Computational Neighborhood at UCSB



Mr. Rogers

Oct 26, 2023



Introduction



- Research Computing
 - Specific scientific computing project you are interested in (or just to experiment)
 - Wondering what resources are available and *where* to do your research, *who* could provide support, and *what* costs, if any, are associated with your project
 - UC Santa Barbara is both a silo'd place and a collaborative place
- In this workshop, you will hopefully learn something about:
 - What computational and data resources are provided at UCSB (that we know about)?
 - A bit about other resources available across campus and who/how to contact about them.
 - Communities across UC Santa Barbara that might be helpful in computational research

If you remember 1 thing from this presentation, it should be

mtc@ucsb.edu or help@grit.ucsb.edu

(since he didn't make the Google group / functional account / annoying ticket system / Slack Channel)



Urban Planning and Public Safety

(In a Research IT Neighborhood Sense)

NOC! & SOC!

NOC - Network Operations Center

Responsible for Interbuilding Network and Internet Access

SOC - Security Operations Center

Responsible for IT Security (campus-wide) & Threat Management & SCRE
(Secure Compute Research Environment)

<http://status.ucsb.edu>

NOC & Local IT Services provide network access to
the wallport (and Wi-Fi - not really research, just FYI)



Storing your Data (or... this is the :(slide)

- Box.com & Google Drive - best for reduced data, documents/papers, (<50GB?)
- Local IT Group Data Shares (LSCG , GRIT, departmental storage)
- AWS - many tiers of storage, some exceptionally inexpensive
info@cloud.ucsb.edu - for more information

Archiving your Data

- Library (<https://www.library.ucsb.edu/research-data-services/repository>)
Dryad - Dryad provides UCSB researchers with user-friendly tools to upload and describe their data for access and reuse and to create citations and persistent identifiers for each dataset.



Research Data Services (RDS)



- Comply with funder mandates (i.e., data management planning and implementation)
- Consultation/assistance on data management (e.g., data documentation, data wrangling, licensing)
- Archiving/publishing and getting credit for data
- Curating datasets (FAIR compliant)
- Training on computational tools to support open science and reproducibility

Contact: rds@library.ucsb.edu



Share/Preserve your data: Data Repositories

- You can't lose data that's in a public data repository
- Support understanding, reanalysis and reuse
- Demonstrate scientific rigor
- Facilitate discovery
- Attract more citations

=> *RDS can help!*

Source:

Morton, L. (October, 18 2021). What a difference a data repository makes: Six ways depositing data maximizes the impact of your science. *Plos Blogs*. <https://theplosblog.plos.org/2021/10/data-repository/>

Research Data Services Data Literacy Series, n.05/2021

How to choose a Data Repository?

Data Repositories (DRs) are large database infrastructures that gather, organize, manage, preserve and make datasets available for research. In this issue we describe the repository types, where you can find them and things you should consider when choosing a DR either to deposit or to find data for reuse.

TYPES & EXAMPLES

General



Includes multidisciplinary data from various fields

Domain-specific



Focuses on a particular discipline or topic area

Search for over 2,600 repositories by discipline, data type, and more on: re3data.org 

REGISTRY OF RESEARCH DATA REPOSITORIES

10 THINGS TO CONSIDER WHEN CHOOSING A DATA REPOSITORY

Does the Repository...

1. Have been recommended by your institution, publishers and/or funding agencies?
2. Assign persistent and unique identifiers to datasets?
3. Provide a landing page for datasets, with metadata that helps others find data, makes it understandable, visible, and promotes the reuse of the data?
4. Track how the data have been used by providing access and download statistics?
5. Follow a sustainable model and is certified as trustworthy, with an explicit commitment to preserve and make data available for the long term?
6. Have capabilities and follow standards that match your particular data needs (e.g. formats accepted, metadata standards, size capacity)?
7. Offer clear terms and conditions that meet legal requirements (e.g. data protection and restrictions) and allow reuse without unnecessary licensing conditions?
8. Provide guidance on how to cite the data that has been deposited?
9. Offer curation services and review data and documentation prior to data publication?
10. Charge for its services?

Still unsure? Contact us: rds@library.ucsb.edu

We can assist you finding and selecting the most appropriate data repository for your research.

UC SANTA BARBARA
Library

www.library.ucsb.edu   

Data Sources

- Brandwatch's social media data library consists of hundreds of billions of posts from content providers including (<https://www.library.ucsb.edu/dreamlab/brandwatch>):

Twitter (formerly known as X): Full Twitter Firehose through a direct partnership with Twitter beginning July 2010, with some content back to July 2009.

Tumblr: Full Tumblr Firehose through a direct partnership with Tumblr beginning January 2015.

Blogs, for example blogspot.com.

Forums, for example reddit.com.

Instagram Hashtags

YouTube: Content from video descriptions and comments.

- UCSB Cylinder Audio Archive (<https://www.library.ucsb.edu/special-collections/performing-arts/cylinders>)
- Just two, as there are too many to list.



Computing on Data (besides your local laptop or workstation)

- UCSB Center for Scientific Computing (CSC) HPC clusters
 - Access to all UCSB entities, Pod/Knot (free) and condo (PI) clusters.
- Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS)
 - National HPC resources funded by NSF. Free*
- NRP Nautilus Cluster (Consumer GPUs - Kubernetes)
 - National cloud computing resource for accelerating machine learning on the GPUs. Free*
- Aristotle Cloud (LSIT)
 - UCSB local cloud resource, e.g. [Jupyter hub](#)
- Secure Compute Research Environment (SCRE)
- Other discipline specific UCSB resources
 - NCEAS, GRIT, LSCG, ECI, your local department
- Commercial Cloud Computing Resources:
 - AWS, Microsoft Azure, Google Cloud Platform



Computing on Data (besides your local laptop or workstation)



Why use these resources?

- Specialized hardware and software (GPUs, tons of RAM, large storage)
- Long Running Computation (days, weeks)
- Multiple Computations Simultaneously (Monte Carlo simulations)

Limitations

- Operating Systems (except for workstations [and maybe AWS nowadays], most are Linux)
- Learning Linux, learning cloud platforms, how to move data, best practices, queuing systems - all take time
- Graphics / Visualization can be difficult
- For free resources, tragedy of the commons

Campus Cloud Landing Zone (LZ)

- for **AWS**, Azure and GCP

LZ for **AWS** = curated environment

- UC-wide Enterprise Discount (EDP)
- S3 Storage Discount
- Enterprise Support Program
- Security & Compliance Guardrails
- Network Connectivity to Campus

Want a Campus Cloud Account?

- Training, Budget Estimate & create PO

**UC Regents Approved
Cloud Providers**

“First and Best in AWS”

<http://Docs.cloud.ucsb.edu>

info@cloud.ucsb.edu



College of Letters & Science

Computational Research Support

L&S Cloud Infrastructure:

Current:

- IaaS: Aristotle Eucalyptus cluster (AWS-compatible)
Send inquiries to <help@lsit.ucsb.edu>
- JupyterHub: deploying environments with both Python and R Studio
[<https://help.lsit.ucsb.edu/hc/en-us/categories/360005255312-Jupyter>]

Coming soon:

- Kubernetes on-premise workloads

Project consulting for L&S research community

(We'll gladly try to help everyone at UCSB get to the best resources) help@lsit.ucsb.edu



GRIT (<https://grit.ucsb.edu/>)

GRIT provides information technology services and technical support for ERI, MSI and ISBER Organized Research Units (ORUs), the Geography department, and the CCBER and NRS Centers as well as other campus organizations.

Support of Research Computing is the primary purpose of GRIT. We are a creative "Can Do" shop and offer a wide range of services and solutions for your computational research and data needs. Data Storage, Scientific Computing, & Custom Solutions. help@grit.ucsb.edu

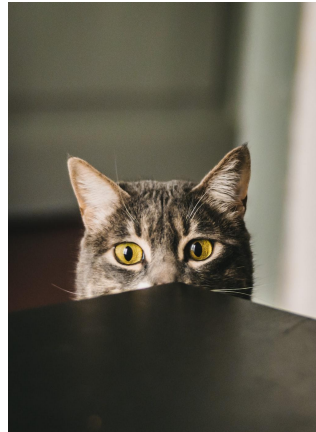


Engineering

Students and Faculty in Engineering should work with their department to figure out the best resource for research IT, as it varies a bit by department and grant. Research IT support is only available to COE researchers. Many COE research groups have substantial local resources - compute, storage, research group support communities from interdisciplinary grants, etc.

Engineering Computing Infrastructure

- Website: eci.ucsb.edu
- Support email: help@engineering.ucsb.edu



Life Sciences Computing Group (LSCG)

- LSCG provides support for researchers in or with a connection to the following departments and units: MCDB, EEMB, Chemistry, Psychology, Earth Science, BioEngineering, NRI, & other smaller associated groups.
- We provide full-stack service covering all IT needs.
- In Research IT specifically, we manage software licenses, group resources, data storage, IoT hardware, instrument controllers, and websites for our users.
- Advice on purchase of computational resources, and administer and manage them once they arrive.
- Represent our users with campus resources, and can setup Google Groups, Box storage, and other solutions based on individual or group needs.

Requests for information, consultation or assistance: help@lscg.ucsb.edu ,
<https://www.lscg.ucsb.edu>



Center for Scientific Computing (CSC)

What we are (<http://csc.cnsi.ucsb.edu>):

- Local (Linux) HPC clusters and expertise with national supercomputing centers leveraging CNSI, MRL, and ITS resources to enable researchers to focus on the research project/education. Some of our compute infrastructure is available to any entity on campus.

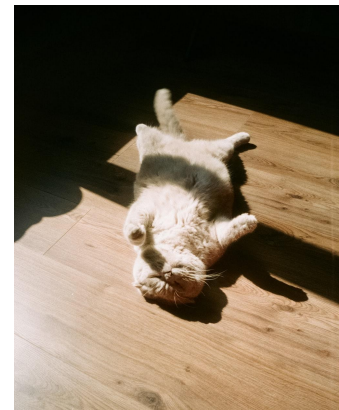
Support Capabilities

- CSC/Grants provide the computational infrastructure.
- Provide a large amount of (temporary, please!) storage for processing data.
- We provide user support, assistance with use of resource, installation of applications, and training/workshops.
- **We work with your local IT staff to provide help.**



HPC Resources of Useful Information

- CSC Software Documentation
 - <https://csc.cnsi.ucsb.edu/docs>
- National HPC resources
 - ACCESS: <https://access-ci.org/>
 - San Diego Supercomputer Center: <https://www.sdsc.edu/>
 - NRP Nautilus: <https://portal.nrp-nautilus.io/>
- Transitioning from XSEDE to ACCESS by using Globus
 - <https://www.globus.org/advance-to-access>
- UCSB Aristotle Cloud (LSIT):
 - <https://www.aristotle.ucsb.edu/> and <https://help.lsit.ucsb.edu/hc/en-us/categories/360005255312-Jupyter>
- UCSB Campus Cloud Information:
 - <https://www.it.ucsb.edu/explore-services/ucsb-campus-cloud>
 - <https://docs.cloud.ucsb.edu/>
- More information, go to <https://csc.cnsi.ucsb.edu/resources>



Communities

GS^3 [Graduate Simulation Seminar Series]- <https://www.facebook.com/gscubed/> (mostly seminars)

Machine Learning - ml@ece.ucsb.edu

Data Science - <https://datascienceucsb.org/> (Focusing on R, Python, and SQL)

Bionews - bionews@lifesci.ucsb.edu (Google Group - only partly research)

Computational Linguistics - ucsb-ceiling@googlegroups.com

Cloud Impact Hub Group - <https://chat.google.com/room/AAAAN8CAkZY?cls=7>

Center for Black Study Research (Data Science) - <https://cbsr.ucsb.edu/research-projects/data-science>

Social Demography Lab - <https://broomcenter.ucsb.edu/research/social-demography-lab>

Center for Information Technology and Society - <https://www.cits.ucsb.edu/>

DreamLab (Collaboratory) Library - <https://www.library.ucsb.edu/dreamlab>

Sciurus griseus



And here's the thing- don't see something for you? Ask, and if we can't find a community resource, create one!

Programming

- Library - Software Carpentry Workshops (<https://ucsbcarpentry.github.io/>) - R, Python, Github, and more.
- VSCode and GitHub Copilot + Chat-GPT4 (3.5 works, but 4 is currently great, though subscription based)
- NCEAS - CoreR - <https://www.nceas.ucsb.edu/learning-hub/core-r> & other NCEAS communities <https://www.nceas.ucsb.edu/learning-hub>
- I'd imagine a few academic departments have programming classes - like Computer Science, seems like something they'd do. I'd also imagine that not all are available to every researcher (plus, you'll probably get a grade :().



Possible Future Idea

- Domain Specialist Research Computing and Data Consultants
Grads or PostDocs, Paid (no one exactly knows how yet), Community Builders
- Berkeley's ad: (<https://research-it.berkeley.edu/news/hiring-research-data-and-computing-domain-consultants>)

Hiring for Research Data and Computing Domain Consultants!

Position overview

Get paid to develop your skills in research data and computing!

UC Berkeley's Research IT unit is hiring graduate student Domain Consultants for flexible, 15% to 25% (6-10 hours/week) appointments. We are seeking diverse and curious candidates from across scholarly disciplines who are enthusiastic about supporting campus research! Research IT supports a broad range of scholarly disciplines and we work to understand their research and facilitate access to computing solutions that best support them.

What you will do

Working with Research IT staff and other domain consultants, you will:

- Offer 1:1 research support to faculty, postdocs, graduate students, and research staff
- Help onboard campus researchers to computing and data solutions
- Develop and contribute to documentation on research data topics and computing solutions
- Identify use cases to support service development
- Participate in (research) instruction and outreach to disciplines across campus

July 31, 2023



Acknowledgements

- If any group helps you extensively with your research, please acknowledge them.
- In the past 6 years, the collaboration between MRL, CNSI, & ITS has produced 500+ peer reviewed publications. Every Research IT group benefits from being recognized.

← → ↺ scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=DMR-1720256+CNS-1725797&btnG=

Google Scholar DMR-1720256 CNS-1725797

Articles About 505 results (0.06 sec)

Any time
Since 2023
Since 2022
Since 2019
Custom range...

Sort by relevance
Sort by date

Any type

(Invited) Investigating Polymorphism in $\text{Na}_2\text{Fe}_2\text{F}_7$ and Its Effect on Electrochemical Properties
EE Foley, VC Wu, W Jin, W Cui... - Electrochemical ..., 2023 - iopscience.iop.org
... This work made use of the shared facilities of the UC Santa Barbara MRSEC (DMR 1720256) and Center for Scientific Computing (CNS 1725797, DMR 1720256). EE Foley and VC Wu ...
☆ Save ↗ Cite

Paying for Integers
J Cross, G Zhang - ... Center for Applied Economic Research Working ..., 2022 - papers.ssrn.com
... Use was made of computational facilities purchased with funds from the National Science Foundation (CNS1725797) and administered by the Center for Scientific Computing (CSC)



Introductions

Paul Weakliem, PhD

Co-Director

Center for Scientific Computing &
California Nanosystems Institute

Eling 3231

weakliem@cnsi.ucsb.edu

Ted Cabeen

Director

Life Sciences Computing Group

LSB 2306

ted@ucsb.edu

Fuzzy Rogers

MRL Computational Facility
Center for Scientific Computing &
Materials Research Laboratory

MRL 2066B

fuz@ucsb.edu

Michael Colee

Director

GRIT

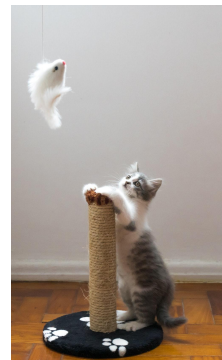
mtc@ucsb.edu

Yu-Chieh “Jay” Chi, PhD

Research Computing Consultant
Center for Scientific Computing &
Enterprise Technology Services

Elings 3229

jaychi@ucsb.edu





Data + Research + Exploration + Access + Methods

Hands-on workshops on topics in research computing: Python, R, Git, Bash

Analytical Software & Tools: ArcGIS Pro, NVivo, MAXQDA, SPSS, Qualtrics

Virtual Reality Studio: Vive, Rift, Hololens headsets, and 3D modeling workstations

Specialized Datasets: geospatial, text corpora, social science data, and satellite imagery from Planet Labs.



Machine learning workshop, summer 2022

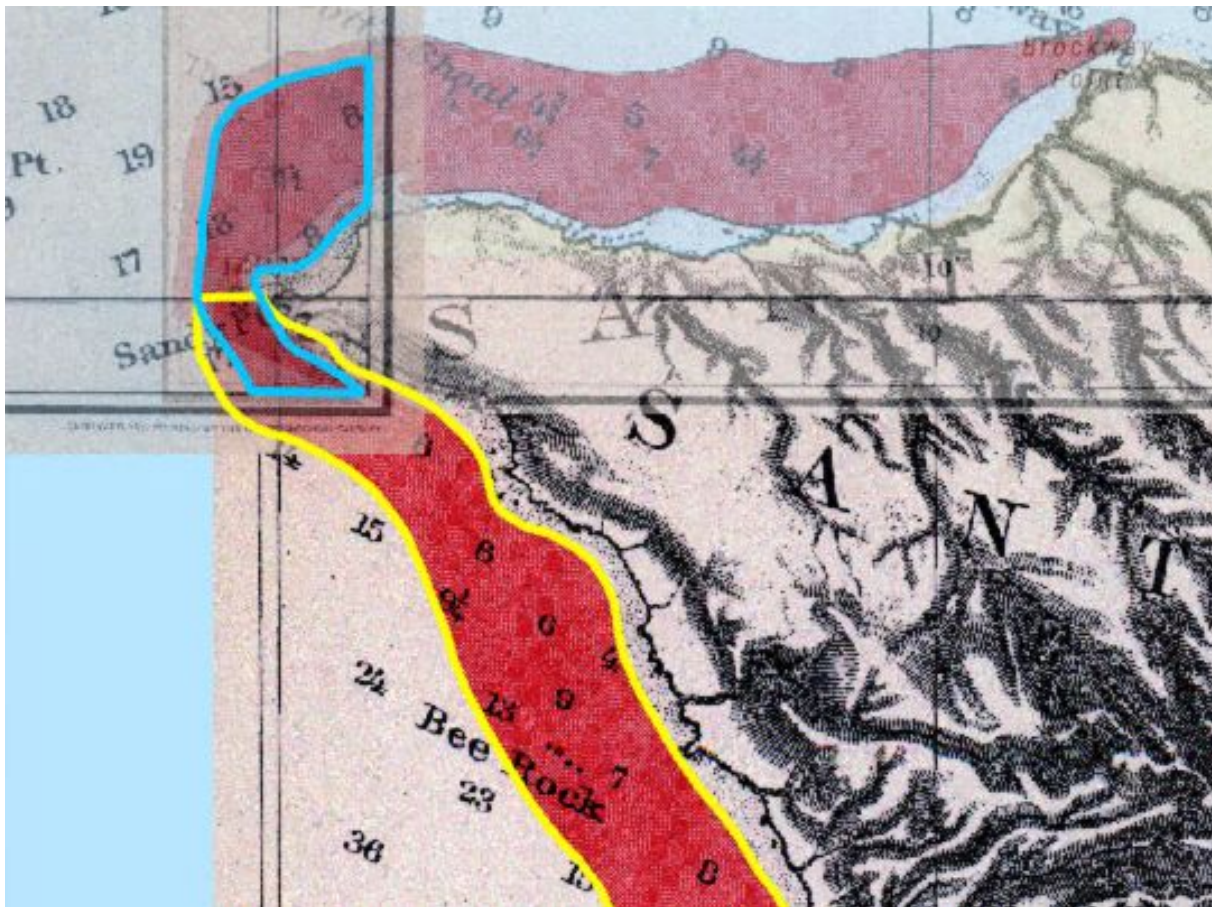
dreamlab@library.ucsb.edu
<https://www.library.ucsb.edu/dreamlab>

GeoSpatial

Esri ArcGIS Site License offers
ArcGIS Online, Living Atlas,
ArcGIS Pro and Enterprise

Extensive data collection
available in the DREAM Lab. ~
500 titles on disk. 3500 data
DVD's

Planet Labs offers a daily
satellite snapshot of Earth back
to 2016



Kelp Groves of the Pacific Coast and Islands c. 1912. Digital vector data created by DREAM Lab students from historic USGS scanned maps



Fall 2023 Workshop Schedule

All Workshops 9:30am - 12:00pm, UCSB Library (Room 1312)

Carpentry Workshop: Introduction to Version Control with Git & GitHub

October 23 -24

Introduction to ESRI Geospatial

November 9th

Carpentry Workshop: Introduction to Data Analysis and Visualization in Python

November 14 -16

Register at: <https://ucsbcarpentry.github.io>

What is the Center for Scientific Computing (CSC)



- Our vision is to provide the computational infrastructure to support multidisciplinary computational research at UCSB.
- Provide computing support and consultation to researchers, including students, post-doctor, and faculties, for software of large-scale scientific computing on the HPC.
- Promote the effective use of the research computing resources of the regional or national area, and scientific software application through training and education, and consultation.
- Directly provide computing support to researchers and assist grant proposals by providing expertise in the assessment of hardware and software requirements.
- Collaborate with other Research IT partners to facilitate the development of research computing/data at UCSB.
- Search for other research computing information from other Universities or Supercomputer centers and bring the workshop topic and present it to UCSB.