

# Vanessa Sochat

*vsoch.github.io*

## EDUCATION

---

### **PhD, Biomedical Informatics:** *Stanford University, Stanford CA*

Sept 2011 – Aug 2016

- Stanford Graduate Student Fellowship (Albion Walter Fellow)
- Microsoft Graduate Women's Scholar (2012)
- National Science Foundation Graduate Fellowship

### **BA in Psychology and Neuroscience:** *Duke University, Durham NC*

Aug 2005 – June 2009

- Magna Cum Laude, Dean's List, Member of Psi Chi, the National Honor Society in Psychology

## EXPERIENCE

---

### **Software Engineer, Research Computing:** *Stanford University, Stanford CA*

Sept 2016 – present

- Systems engineer and architecture, infrastructure and standards development, and specialized technical consultation to better ensure reproducibility of scientific computational analyses and workflows
- Implemented [Snakemake Google Life Sciences](#) executor backend using several Google Cloud APIs
- Host and producer of [Research Software Engineer Stories podcast](#)
- Designed the [Scientific Filesystem](#) for organization and discovery of scientific applications
- Developer of [Singularity Hub](#), [Singularity Registry](#), and original developer for [Singularity](#) containers
- Lead of open source project [The Experiment Factory](#) for reproducible behavioral experimentation
- Implemented and serviced image processing pipelines for the School of Medicine.
- Complete list of work available at <https://vsoch.github.io/work>

### **PhD Candidate, Poldrack Lab:** *Stanford University, Stanford CA*

June 2011 – Aug 2016

- Designed and developed a Dockerized infrastructure, [expfactory.org](#), to deploy web-based experiments
- Conceptualized and implemented open source software, [Wordfish](#), for generating custom NLP pipelines
- Created an [interactive, reproducible workflow](#) to for genomic, behavioral, and brain imaging analyses
- Identified [optimal parameters](#) for comparison of statistical brain maps using classification framework
- Imagined and created web viewers for brains using [nodeJS](#), a [neuroimaging data model](#), and [FileReader](#)
- Built clinical [web application](#) to explore anatomical and genomic features associated with brain tumors
- Built model and [database](#) to classify artifact in functional MRI using regularized logistic regression
- Created complicated analysis pipelines in a HPC environment to analyze thousands of brain images

## **Data Technician, Laboratory of Neurogenetics: *Duke University, Durham NC***

May 2009 – May 2011

- Coded and deployed image processing pipelines in HPC environment using python, bash, and Matlab
- Wrote custom tools to check the quality of brain images, organize data, and interact with participants
- Responsible for creating and administering a battery with over 30 cognitive paradigms using Qualtrics

## **Founder, Goggles Optional Podcast: *Stanford University, Stanford CA***

Nov 2013 – present

- Developed and currently maintain infrastructure for a weekly science podcast with over 50K downloads
- Weekly responsibility to generate episode content, update databases, and publish

## **Student Director, Informatics Concentration for MD Students: *Stanford CA***

May 2013 – May 2015

- Organized quarterly sessions for approximately 30 medical students interested in informatics
- Set up social media groups and advertising for MD student recruitment

## **Teaching Assistant, Biomedical Image Analysis and Interpretation: *Stanford CA***

Jan 2013 – May 2014

- Created new course content for 10 lectures, including interactive slides and class handouts
- Single handedly developed two new projects, including a database of “cookie tumor” images
- Taught weekly section meetings, and gave two full lectures on machine learning and neuroinformatics

## **SKILLS AND QUALIFICATIONS**

---

### **Computer Experience**

<i>Languages:</i>	Python, bash, R, JavaScript, Matlab, HTML/CSS, php, GoLang
<i>Databases:</i>	MySQL, PostgreSQL, neo4j, couchdb, Big Query, sqlite3
<i>Infrastructure:</i>	Docker, Singularity, VirtualBox, Vagrant
<i>Visualization:</i>	D3, canvas, Shiny (R), Photoshop, Illustrator, Maya, Blender

### **Data Analysis**

<i>High Performance Computing:</i>	SLURM, SGE
<i>Data Structures</i>	JSON, xml/RDF, yaml

### **Web Development**

<i>Frameworks:</i>	Django, Jekyll, Flask, Wordpress, nginx, uWSGI
<i>Cloud Technology</i>	Google Cloud, AWS (EC2, RDS, S3)
<i>Continuous Integration</i>	CircleCI, Travis, GitHub Actions
<i>Version Control</i>	GitHub, Bitbucket, Gogs

## SELECTED PUBLICATIONS

---

Mölder F, Jablonski KP, Letcher B, Hall M, Tomkins-Tinch C, **Sochat V**, et al. Sustainable data analysis with Snakemake [version 1; peer review: awaiting peer review]. F1000Research 2021, 10:33 (<https://doi.org/10.12688/f1000research.29032.1>)

**Sochat, V.** (2020). AskCI Server: Collaborative knowledge base. In Practice and Experience in Advanced Research Computing (PEARC '20). Association for Computing Machinery, New York, NY, USA, 514–517. DOI:<https://doi.org/10.1145/3311790.3399616>

**Sochat, V.,** (2020). GridTest: testing and metrics collection for Python. Journal of Open Source Software, 5(51), 2284, <https://doi.org/10.21105/joss.02284>

**Sochat, V** (2019). Singularity Compose: Orchestration for Singularity Instances. Journal of Open Source Software, 4(40), 1578, <https://doi.org/10.21105/joss.01578>

**Sochat, V** (2019). Container Tree: Software to Model Container Filesystems, Packages, and Inheritance. Journal of Open Source Software, 4(37), 1418, <https://doi.org/10.21105/joss.01418>

**Sochat, V** (2019). WatchMe: Software for Reproducible Monitoring and Data Collection. Journal of Open Source Software, 4(37), 1388, <https://doi.org/10.21105/joss.01388>

**Sochat, V** (2018). Containershare: Open Source Registry to build, test, deploy with CircleCI . Journal of Open Source Software, 3(28), 878, <https://doi.org/10.21105/joss.00878>

**Sochat, V** (2018). HelpMe Command Line Helper Utility . Journal of Open Source Software, 3(26), 775, <https://doi.org/10.21105/joss.00775>

**Sochat V**, (2018). The Scientific Filesystem. GigaScience, giy023, <https://doi.org/10.1093/gigascience/giy023>

**Sochat V**, (2018). The Experiment Factory: Reproducible Experiment Containers. Journal of Open Source Software, 3(22), 521, <https://doi.org/10.21105/joss.00521>

**Sochat V**, Prybol CJ, Kurtzer GM (2017) Enhancing reproducibility in scientific computing: Metrics and registry for Singularity containers. PLoS ONE 12(11): e0188511. <https://doi.org/10.1371/journal.pone.0188511>

**Sochat V**, (2017), Singularity Registry: Open Source Registry for Singularity Images, Journal of Open Source Software, 2(18), 426, doi:10.21105/joss.00426

Kurtzer GM, **Sochat V**, Bauer MW (2017) Singularity: Scientific containers for mobility of compute. PLoS ONE 12(5): e0177459.

**Sochat V**, Eisenberg IW, Enkavi AZ, Li J, Bissett PG and Poldrack RA. The Experiment Factory: standardizing

Sochat

behavioral experiments. *Front. Psychol.* 2016.

Durnez J, Degryse J, Moerkerke B, Seurinck R, **Sochat V**, Poldrack R, Nichols T. Power and sample size calculations for fMRI studies based on the prevalence of active peaks. *bioRxiv*, 2016.

**Sochat V**, Gorgolewski KJ, Koyejo O, Durnez J, Poldrack RA. Effects of thresholding on correlation-based image similarity metrics. *Frontiers in Neuroscience*. 2015.

**Sochat V**, AuthorSynth: a collaboration network and behaviorally-based visualization tool of activation reports from the neuroscience literature. *Frontiers in Neuroinformatics*. 2015.

Poldrack, R, Laumann T, Koyejo O, Gregory B, Hover A, Chen MY, Gorgolewski KJ, Luci J, Joo SJ, Boyd R, Hunicke-Smith S, Simpson Z, Caven T, **Sochat V**, Shine J, et al. "Long-Term Neural, Behavioral, and Physiological Phenotyping of a Single Human: The MyConnectome Project" *Nature Communications*. 2015.

**Sochat V**, Supekar K, Bustillo J, Calhoun V, Turner JA, et al. A Robust Classifier to Distinguish Noise from fMRI Independent Components. *PLoS ONE*. 2014.

S. Finlayson, **V. Sochat**, L. Szabo, L. Yancy Jr. A Rapid Learning System for Personalized Glioblastoma Treatment Planning. Abstract presentation at the AMIA Annual Symposium, Washington DC, USA. 2013.

## SELECTED TALKS AND PRESENTATIONS

---

**Sochat, V** (2020, September) “Singularity Container on Google Cloud Platform (GCP) Tutorial” Co-host for Google HPC Days, [https://cloudonair.withgoogle.com/events/singularity\\_containers\\_on\\_gcp\\_tutorial](https://cloudonair.withgoogle.com/events/singularity_containers_on_gcp_tutorial).

**Sochat, V** (2020, September) “Extensions to Schema.org for structured, semantic & executable research documents” Invited panelist, Stencila Community Webinar  
<https://stenci.la/blog/2020-09-20-register-now-for-stencila-community-call-thur-24-sept-2020/>.

**Sochat, V** (2020, September) “Research Software Directories, What, Why, and How?” Series of Online Research Software Events (SORSE) 2020 Talks and Discussion,  
<https://sorse.github.io/programme/discussions/event-013/>.

**Sochat, V** (2020, July) “The Research Software Encyclopedia” Invited talk, Research Software Engineers Community Workshop, 2020, PEARC20, <https://us-rse.org/events/2020/2020-07-pearc20>.

**Sochat, V** (2020, July) “The Singularity Executor: A Contributor Overflow Exception” Invited talk, Apache Airflow Summit 2020, <https://airflowsummit.org/speakers/vanessa-sochat/>.

**Sochat V**, (2020, June) “Singularity Containers” Invited talk, Dataverse Community Meeting,  
<https://projects.iq.harvard.edu/dcm2020/people/vanessa-sochat>.

**Sochat V**, (2020, January 10) “Research Software Engineering: A Future at Stanford” Keynote: Campus IT Plan Research Summit, Stanford CA, <https://itcommunity.stanford.edu/events/campus-it-plan-research-summit>.

**Sochat V**, (2019, December 19) “Research Software Engineers: A New Career” Lightning Talk: IT Unconference, Stanford CA, <https://itcommunity.stanford.edu/unconference>.

**Sochat V**, (2018, April 4) “The Scientific Filesystem” Invited Speaker: Containers in HPC Symposium at UCAR, Boulder CO, <https://sea.ucar.edu/conference/2018/containers>.

**Sochat V**, (2018, March 7) “Introduction to Singularity” Invited Speaker: CyVerse Container Camp: Container Technology for Scientific Research, University of Arizona, Tuscon AZ.

**Sochat V**, (2017, July 11). “Reproducibility and Containers: The Perfect Sandwich” Invited Speaker: Practice & Experience in Advanced Research Computing, New Orleans LA. <https://www.pearc17.pearc.org/speakers>

**Sochat V**, (2017, February). “Singularity Containers for Scientific Compute” Talk Stanford Genomics Cluster User Group, Stanford CA, USA.

**Sochat V**, (2015, October). “Building Tools for Neuroimaging: the intersection of high performance computing, web technology, and fun in graduate school.”, Talk for Research Computing Group, Stanford CA, USA.

**Sochat V**, (2015, March). “Brain Maps Like Mine content-aware image comparison and retrieval for interactive visualization and meta-analysis of brain statistical maps”, Research in Progress Talk, Stanford CA, USA.

**Sochat V**, (2014, June). “Introduction to Machine Learning,” SIMR Summer Research Program, Stanford CA, USA.

**Sochat V**, (2014, May). “Machine Learning for Images,” Biomedical Imaging Analysis & Interpretation Lecture, Stanford CA, USA.

**Sochat V**, (2013, May). “Neuroinformatics,” Biomedical Imaging Analysis and Interpretation Lecture, Stanford CA, USA.