

# Carnegie Mellon's first Open Science Symposium – Themes about research data and their reuse

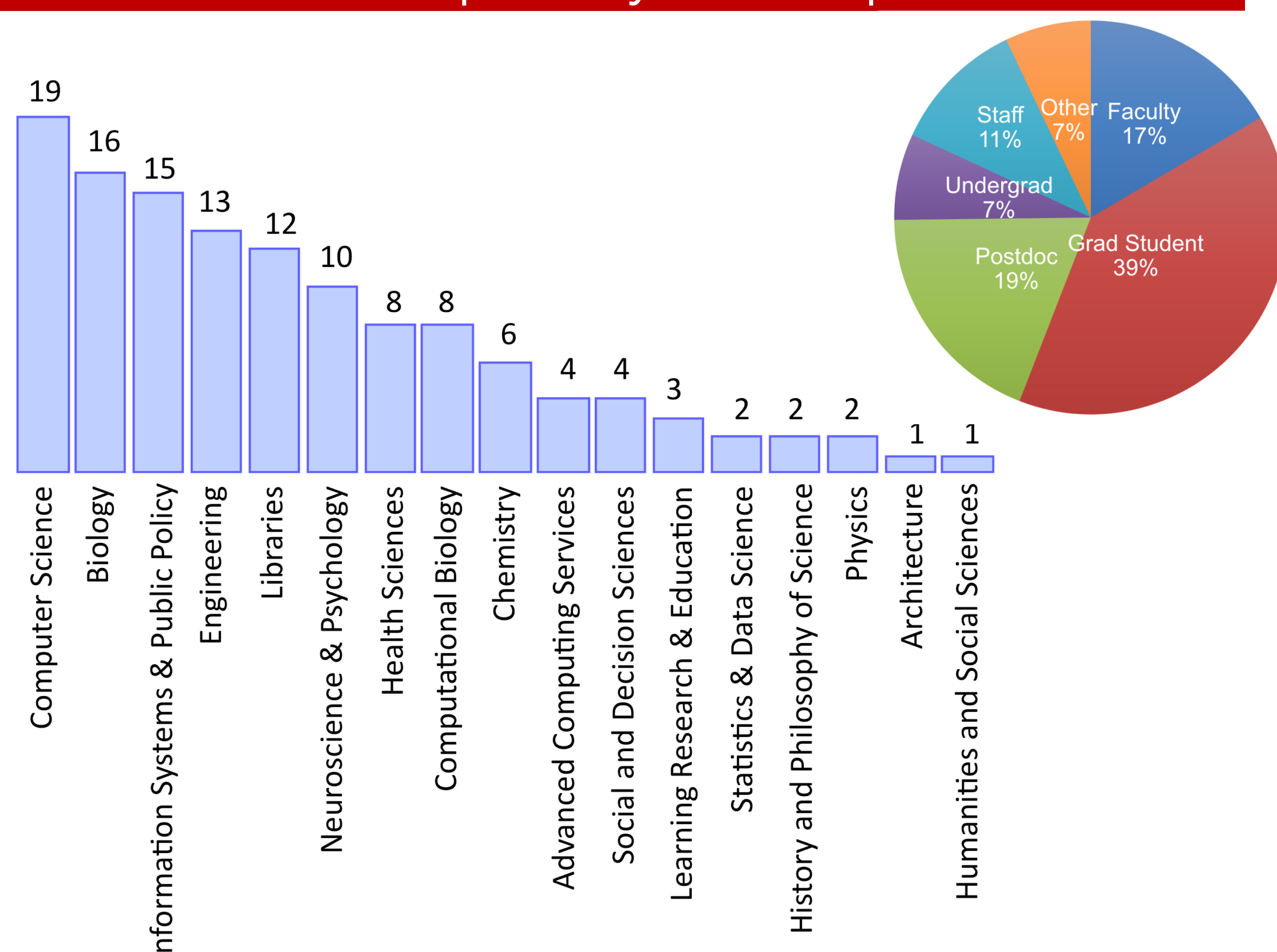
Huajin Wang, Melanie Gainey, Ana Van Gulick  
University Libraries, Carnegie Mellon University, Pittsburgh, PA, USA



## Background

In October 2018, Carnegie Mellon University (Pittsburgh, PA, USA) hosted its first Open Science Symposium: <https://events.mcs.cmu.edu/oss2018/>. The two-day event was a collaboration between data specialists in the University Libraries and life sciences faculty and was funded by a grant for life sciences from the DSF Charitable Foundation. The goal of the event is to build awareness and support for the adoption of open research practices, and encourage innovative ideas about data sharing in the life sciences and related disciplines. Presentations from the event have been documented on Open Science Framework: <https://osf.io/54gue/>.

## Interdisciplinary Participation



## Sessions

DAY 1

### Invited talks + Panels (18 speakers):

- Open Science in Research
- Open Data and Reproducibility
- Open Tools and Platforms
- Open Access

### Scientific Speed Dating + Reception

DAY 2

### Workshops:

- Code Ocean
- Bioconductor
- KiltHub Deposit-a-thon
- BenchSci
- Pittsburgh Supercomputing Center
- ENCODE DCC
- protocols.io

## Emerging Hot Topics

- Open hardware and software
- Large scale team science / observatory science
- Open framework for citizen science
- National-wide data collection and curation:
  - NCI cancer research commons
  - NCBI Short Read Archive
  - ENCODE DCC
- Different flavors of reproducibility
- Reproducible computing on HPC
- Evolving platforms by publishers:
  - eLife reproducible articles
- Pre-printing: Pros and cons

## Recurring Themes

Incentives and standards for data sharing need to come from the ground-up within the disciplines.

New collaborative methods for data collection and analysis are reshaping the way that science is done.

Open data ≠ reusable data. Reproducibility and reusability make data valuable, not just open.

“There were microscopes that were too expensive for us ...why don’t we make it ourselves and make it open source?”  
– Miniscope Project. Mount Sinai.

“...make tools more accessible for others, and others will contribute back”  
– Open Ephys Project. Allen Institute.

“...institutions jointed force and use the observatories approach to collect and analyze brain fMRI data (like what astrophysicist have been doing)” – Brain Observatories. University of Washington.

“With consortia effort (for DNA data collection and curation)... the goal is to find data as easy as finding shoes on Zappos.” – ENCODE Data Coordination Center.

“We need a lot of labeled data... Swipe left for good image, swipe right for bad image.” – Braindr, a citizen science platform. University of Washington.

“Till we get to a point of how to measure the value and reuse of data, we won’t have incentives in place” – Panel Discussion.

“Grabbing data via point and click is not reproducible” “With a few lines of code, get the data reproducibly (via API). If data changes, just have to re-run the code” – NCI Data Commons.

“Open access ≠ useful resources... implementation is not FAIR” – ENCODE CCD

“Science always moves faster than evaluation processes; it’s going to change when some of us are on the other side.”  
– Panel Discussion.