

# Supporting Big Data Research: Recommendations from the Ithaka S + R Projects at UCB and UCSD

---

**2021 UC Libraries Forum  
October 28, 2021**

**UCB:**  
Brian Quigley  
Ann Glusker  
Erin Foster

**UCSD:**  
Reid Otsuji  
David Minor  
Stephanie Labou

Ithaka S+R is part of ITHAKA, a not-for-profit organization helping the academic community use digital technologies to preserve the scholarly record and to advance research and teaching in sustainable ways.

# Introduction

“A partnership with academic libraries to examine the research support needs of faculty using “big data” and data science methodologies.”

- Ithaka S+R “Supporting Big Data Research”
- Brief synopsis of the Ithaka S+R research study process
- Independent UCB and UCSD research reports produced:
  - Common findings & important differences
  - Potential cross-campus collaboration opportunities

# Common Findings

- Research data processes
  - Commitment to openness
- Infrastructure for big data support
  - Lack of familiarity with campus resources
  - Costs play a major role in decisions around data storage
- Communication channels & promoting services
  - Library partnerships and resource sharing with Research IT
  - Lack of familiarity of resources/services
- Training needs for researchers
  - Foundational skills training
  - Graduate students and Postdocs

# Important Differences

- Distinctive contexts and histories
  - UCSD: presence of SDSC and medical school
  - UCB: launch of renewed Library Data Services Program
- Interest in cloud computing & storage
- Concerns about long-term archiving
- Incentives for data/code sharing and implications for promotion & tenure

*“In terms of resources, I’m a big fan of buying hardware, not renting it. We’ve done stuff on the cloud before, but it always adds like 30%, in time and cost to every project. And as soon as you spend that money, it’s gone.” (UCSD)*

*“We try to make everything as inexpensive as possible and as easy to retrieve as possible. And cloud storage is a simple solution. Without having to pay for the ingress or the egress charges that Amazon or Google or Cloud charges because we are an academic institution...” (UCB)*

# Cross-Campus Collaboration Possibilities

- Formalize cross-campus virtual data training opportunities
- Continue collaborative Love Data Week events and other programming
- Launch a broad-based campaign to promote openness and sharing
- Computing and storage infrastructure? Share expertise and ideas

*“The tsunami is coming. [...] The question is, where in that spectrum do [we] want to be? [...] Do we want to be able to meet the bare minimum complying with big data capabilities? Or do we want to make sure that big data is not an issue? Because the thing is that it’s thrown around in the context that big data is a problem, a buzzword. But how do [we] make that a non-buzzword?”*

*Big data should be just a way of life. How do we get to that point?”*

# Thanks!

UC Berkeley:

Ithaka S+R, UC San Diego Project Team, UC Berkeley interviewees, Salwa Ismail, Anthony Suen, campus partners, and library colleagues

UC San Diego:

Ithaka S+R, UC Berkeley Project Team, UC San Diego interviewees, campus partners, and The Library

# Questions?

Full reports are available here:

UC Berkeley Report:

<https://doi.org/10.25350/B59G6F>

UC San Diego Report:

<https://escholarship.org/uc/item/8kr7p2c0>