





# Journal of eScience Librarianship

putting the pieces together: theory and practice

## Bridging Data Communities: Interoperability through inclusive, cross-institutional collaboration

**Anna Sackmann**, University of California, Berkeley, Berkeley, CA, USA, [asackmann@berkeley.edu](mailto:asackmann@berkeley.edu) 

**Lisa Ngo**, University of California, Berkeley, Berkeley, CA, USA 

**Elliott Smith**, University of California, Berkeley, Berkeley, CA, USA 

**Misha Coleman**, University of California, Berkeley, Berkeley, CA, USA 

### Abstract

**Objectives:** To demonstrate how librarians can use engagement strategies to foster the exchange of knowledge and skills for data analysis and to build bridges between data communities. A second objective is to help student instructors to develop effective live-coding pedagogical practices and to gain practical experience in leading participatory workshop sessions.

**Methods:** Librarians developed a low-barrier introductory peer-to-peer data science workshop series to support students seeking to develop coding, data analysis, and visualization skills, with a focus on Python and SQL. We guided undergraduate peer instructors in participatory live-coding pedagogy, organized practice sessions for instructors, and managed the scheduling, logistics, outreach, and hosting of the workshops.

**Results:** In Fall 2023 sessions in the workshop series were delivered synchronously to over 100 participants, including students from our home institution and more than a dozen community colleges; one workshop was delivered twice—once in English, once in Spanish. Workshop recordings posted online have been viewed over 1000 times.

**Received:** July 18, 2024 **Accepted:** November 14, 2024 **Published:** December 4, 2024

**Keywords:** research data services, data literacy, data science, training, outreach, peer-to-peer instruction, collaboration

**Citation:** Sackmann, Anna, Lisa Ngo, Elliott Smith, and Misha Coleman. 2024. "Bridging Data Communities: Interoperability through inclusive, cross-institutional collaboration." *Journal of eScience Librarianship* 13 (3): e970. <https://doi.org/10.7191/jeslib.970>.

The *Journal of eScience Librarianship* is a peer-reviewed open access journal. © 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited, and new creations are licensed under the identical terms. See <https://creativecommons.org/licenses/by-sa/4.0>.

∞ OPEN ACCESS

## Abstract Continued

**Conclusions:** We successfully identified strategies for building upon existing relationships and strengthening connections among diverse data communities; designing programs and outreach efforts to lower barriers to participation in data science; and fostering a culture of diversity, equity, and inclusion in data science knowledge sharing.

---

## Introduction

As the field of data science has grown in the last two decades so too has the drive to incorporate data literacy into the undergraduate curriculum (National Academy of Sciences, Engineering et al. 2018; Ridsdale et al. 2015). In 2023, the California Department of Education went further and adopted the *Mathematics Framework for California Public Schools: Kindergarten Through Grade Twelve (Mathematics Framework)*, which includes a data science component (California Department of Education 2023), setting an expectation for California students to have some data literacy upon graduation from high school.

On the University of California, Berkeley (UC Berkeley) campus, even prior to data literacy becoming standard in the curriculum, we had anecdotal evidence that students experienced an unspoken expectation for incoming undergraduates to have had some exposure to data science concepts. This was confirmed in a 2019 informal exploratory needs survey of UC Berkeley College of Engineering undergraduates, where we found that not all students had been exposed to data science concepts, and those who wanted to quickly familiarize themselves were not able to find support to do so. These findings align with Wilson's observation that students who are not in computer science often experience an absence of institutional support for novice learners to develop data analysis skills (Wilson 2016). Students specifically articulated a need for support outside of the traditional classroom setting in learning the basics of Python for data analysis and visualization, in some cases prior to enrolling in one of the many credited data science courses offered on campus.

To address the support gap, the UC Berkeley Library developed a sustainable and scalable solution by partnering with the Data Science Discovery Consultants (Discovery Consultants) in the College of Computing, Data Science, and Society (CDSS). The Discovery Consultants are undergraduate students employed by CDSS; their majors include data science and related fields, and they receive training to offer consultation services across a wide range of topics including Python, R, SQL, and Tableau (Brown et al. 2024). In collaboration with the Discovery Consultants, the Library developed an Introduction to Data Analysis workshop series that covers topics such as Python, SQL and Tableau. Launched in 2020, the workshop series grew to reach dozens of departments and programs as well as underrepresented groups on the UC Berkeley campus and California Community Colleges. This paper addresses building the partnership between the Library and CDSS, our train-the-trainer approach with the Discovery Consultants, workshop

content development, and our outreach strategies. For institutions interested in pursuing a similar program, we offer our lessons learned and effective strategies for successful collaborations with undergraduate student partners.

## Discussion

### *Library and CDSS Partnership*

Libraries play a pivotal role in building community and serving as a central hub for discovering and generating new skills and knowledge (Oliver et al. 2019). Data science transcends disciplinary barriers that impacts research and has implications on daily life including public health, navigating climate change, and the structure of our communities (Hudson Vitale, Kennedy, and Ruttenberg 2022). This broad range of applications makes the university library an ideal home for learning and building community around data science, particularly at the introductory level. Academic libraries around the country support data science departments and colleges as they emerge and grow. The University of Arizona Libraries expanded their services by hiring more staff to support computational literacy, geographic information systems (GIS), and reproducible science (Oliver et al. 2019). The University of Minnesota Libraries found that workshops on data ethics were a natural extension of their support of research data and data stewardship (Mani and Cawley 2022). Librarians at Florida State University recognized an opportunity to provide a peer-to-peer STEM data fellowship program, which provided the Library with enhanced data services and offered career development opportunities for undergraduates (Ruhs 2023). At UC Berkeley in 2017, before CDSS became a college, the division partnered with the Library to provide data science peer consulting led by upper division undergraduates. In this instance, the Library provided space, structure, and training to the undergraduates to answer data science related questions. Because the UC Berkeley Library and CDSS had previously established a partnership through data science peer consulting, providing workshops on data analysis for novice learners through a collaboration between the Library and the Discovery Consultants was a natural fit.

### *Outreach*

The workshop series was taught during the fall semester with a period of training and instructor development during the preceding summer term. Each year required iterating, improving, and adapting the previous versions of the workshop series to reflect changes in the Library, CDSS, and student needs. A main goal for the partnership was to create a scalable and sustainable program that develops the undergraduate consultants as instructors. We provided them with pedagogical guidance while they developed workshop content that resonated with novice learners. The series was not restricted to specific attendees; however, the Library gradually moved from targeted outreach in the College of Engineering to include disciplines in the life and health sciences, and ultimately to campus-wide outreach. Initially we followed the exploratory needs survey of College of Engineering students by targeting them for outreach that promoted registration for the workshops. This outreach plan leveraged our existing connections with the Engineering student newsletter and other College of Engineering communication channels.

As the workshop series progressed, librarians from other parts of the Library with departmental liaison duties and expertise joined the Library team. We extended our outreach to the life and health sciences when our Biology and Bioinformatics Librarian joined the team, and we were able to take a broader approach with outreach to undergraduate students when an additional librarian joined who specialized in information literacy instruction. This resulted in attendance from students coming from a wider variety of disciplines.

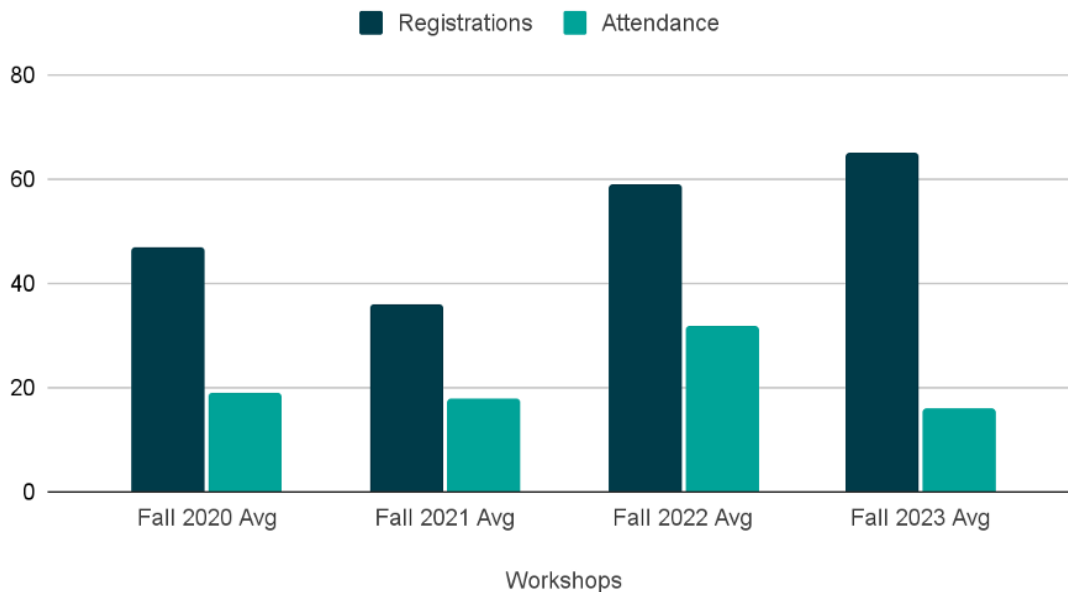
During the workshop series, we learned that CDSS had grown their partnerships and support of students who want to study data science in the California Community College system. Additionally, UC Berkeley established a task force with the goal of becoming an Hispanic Service Institution (HSI), defined as an institution of higher education that has an enrollment of at least 25 percent Hispanic undergraduate full-time-equivalent students, by 2027 (Office of the Law Revision Council, United States House of Representatives 2009). Enrollment indicates that around 22% of our undergraduate students identify as Hispanic/Latino in the 2023-2024 academic year (2024). By contrast, 46% of the students enrolled at California Community Colleges are Hispanic (California Community Colleges 2019). Based on the goal to become an HSI and CDSS's outreach extension to California Community Colleges, we included community colleges in our outreach and offered an introduction workshop in Spanish.

Targeting outreach for departments and student groups (anecdotally) led to better registration and attendance numbers. Taking a more blanket approach relied on people with whom we have less formal connections to read the email or notice the calendar invite. Leveraging personal connections where trust has already been built resulted in the best outreach outcomes.

Because we began the workshop series in 2020 during the COVID pandemic when UC Berkeley was not holding classes in person, most of our workshops were taught over Zoom. Average workshop registration and attendance data increased as outreach to students broadened (figure 1). In Fall 2023, the Discovery Consultants expressed interest in developing in-person teaching skills and decided to teach one introduction workshop in a Library instruction space instead of over Zoom. This workshop was one of the lowest attended of the introduction workshops, which provided us with feedback about our learners' preferred instruction space. The Spanish language workshop also had low attendance; however, as discussed below, it had a high number of asynchronous views over YouTube.

Following each workshop, we captioned the recordings and made them available to anyone who registered through direct links to the recordings in Google Drive. This last year, partially due to YouTube's captioning capabilities, we made the recordings available through a playlist that is viewable by anyone. This made the workshop content more widely available and resulted in a greater number of viewers. As of October 2024, the Introduction to Data Analysis workshop has been viewed on YouTube 948 times, SQL for Data Science has been viewed 301 times and Introducción al análisis de datos con Python has been viewed 770 times.

## Average Workshop Series Registrations and Attendance



**Figure 1:** Average workshop registration and attendance rates.

### *Development of Content*

To develop workshop content, librarians worked with the Discovery Consultants to adapt lessons from open-source curricula from The Carpentries, an international non-profit that develops and hosts workshops teaching data and computational skills. The Carpentries' lessons employ an active learning approach that includes participatory live coding and hands-on exercises. Raj et al. found that live coding makes programming more approachable for novice programmers, helps them learn the process of debugging, and exposes them to good programming practices (2018). They also note that students prefer coding along with the instructor during lessons. To make lessons open for others' use, we presented our adapted lessons in Jupyter Notebooks and made them readily accessible via GitHub.

The core Carpentries values of empowering one another, of valuing all contributions, and fostering inclusive, welcoming learning environments resonated with the Discovery Consultants, and they enthusiastically put those values into practice in their workshops. Throughout our lessons we emphasized that we are all learners, including the instructors. The workshop offerings changed slightly each semester since 2020 based on instructor capacity, participant feedback from previous workshops, and Discovery Consultants' firsthand knowledge of fellow students' needs from their consulting work.

One recent addition to the workshop offerings that exemplified our student instructors' commitment to fostering inclusive learning environments has been the development of a Spanish-language introductory Python workshop. A consultant who is a native Spanish speaker was motivated to develop an environment where native speakers could better connect with the content and with each other. Instead of teaching the Spanish language content from *The Carpentries*, he made significant adaptations to our shortened version of *The Carpentries Python* lesson and provided thoughtful explanations of programming terminology and example problems tailored for a Spanish-speaking audience. This approach enabled him to develop his own teaching style. A Spanish-speaking member of the librarian team supported him in crafting content, rehearsing, and delivering the workshop last fall. Librarians at Northeastern Illinois University found that peer mentors were instrumental to building connections between Latinx students and the Library, and those connections helped serve students more effectively (Green 2011). Developing this workshop also aligned with UC Berkeley's goal of becoming a Hispanic Serving Institution (HSI).

### *Train-the-trainer approach*

This workshop series is a partnership between the Library and the Discovery Consultants in CDSS, using a "train-the-trainer" approach. This model has benefits for everyone involved. Pon-Barry et al. observe that peer instruction in entry level computer science courses can create an effective learning environment that helps combat alienation in introductory courses (2017). They saw better comprehension, better grades, and improved pass rates, particularly for underrepresented students. Peer instructors gain experience designing and leading workshops and in teaching technical skills, and the librarians can reach more students than the Library could as sole service provider.

For this series, librarians were responsible for coaching the Discovery Consultants on how to provide effective, inclusive instruction. We set a time-table and provided training sessions on instructional design and teaching technical skills. Together with the Discovery Consultants we strategized about the content we wanted to offer for the semester's sessions. The Discovery Consultants developed and delivered the workshops. As the student instructors developed their content, we built in time for them to practice delivering it—both in person and recorded—and to receive feedback. The Discovery Consultants delivered the workshops with the librarians providing support as organizers, hosts and helpers.

## **Conclusion**

The data analysis workshop series was successful for four key reasons. First, it addressed needs identified and expressed by students themselves. Second, it leveraged existing sources of coding expertise, the Discovery Consultants, and aligned with their program goals. Third, it employed an effective, tested, open and inclusive curriculum, *The Carpentries*; using this curriculum, the student instructors build their pedagogical and presentation skills under the guidance of librarians. Finally, it met the students where they are by providing multiple learning modes (in person, synchronous Zoom and asynchronous recordings) and convenient

scheduling. The series was designed to meet the needs of diverse participants, including UC Berkeley and community college students as well as Spanish speakers, and to lower barriers to participation in data science.

The greatest challenge in implementing the series was time constraints. Librarians, Discovery Consultants, and the Data Science Discovery program director all have very full schedules. Pedagogical training, outreach, and developing and delivering workshops are time-intensive activities that can present challenges to coordination and planning. To stay on track, good communication, scheduling sufficient time for each stage of the process, and regular check-ins to assess progress have been key. Distributing tasks across the team and recruiting additional librarians ensures that the workshop series is scalable and sustainable.

Organizing and hosting the data analysis series enabled librarians to enhance connections with campus partners and support students from multiple disciplines and career stages. The success of the series demonstrated the value of the Library and helped to fulfill its mission of service. We will continue to assess and adapt our efforts based on feedback, capacity, and partner priorities to meet the data science needs of students at UC Berkeley and beyond.

---

### Competing Interests

The authors declare that they have no competing interests.

### Acknowledgements

The authors would like to acknowledge Kristina Bush, Emily Liu, Arlo Malmberg, and Anthony Suen for their contributions to the partnership.

The content of this article is based on the presentation entitled “Bridging Data Communities: Interoperability through inclusive, cross-institutional collaboration” originally presented at [RDAP Summit 2024](#), available from Open Science Framework: <https://osf.io/yrxhn>.

---

### References

- Brown, C. Taylor, Megan Mehta, Mahathi Ryali, Xiaoran Dong, Iliya Shadfar, Jacqueline Dominquez Davalos, Aaron Culich, and Anthony Suen. 2024. “The Data Science Discovery Program: A Model for Data Science Consulting in Higher Education.” *Stat* 13 (2): e677. <https://doi.org/10.1002/sta4.677>.
- California Community Colleges. 2019. “Frequently Asked Questions.” <https://www.cccco.edu/About-Us/Chancellors-Office/Divisions/Digital-Innovation-and-Infrastructure/research-data-analytics/data-snapshot/student-demographics#:~:text=How%20many%20students%20are%20enrolled,is%20closer%20to%201.8%20million>.
- California Department of Education. 2023. “Mathematics Framework.” <https://www.cde.ca.gov/ci/ma/cf>.

Green, David. 2011. "Supporting the Academic Success of Hispanic Students." In *College Libraries and Student Culture*, 87-108. American Library Association.

Hudson Vitale, Cynthia, Mary Lee Kennedy, and Judy Ruttenberg. 2022. "Advancing Data Science, Data-Intensive Research, and Its Understanding Through Collaboration." In *Handbook of Research on Academic Libraries as Partners in Data Science Ecosystems*, 1-20. IGI Global.  
<https://doi.org/10.4018/978-1-7998-9702-6.ch002>.

National Academy of Sciences, Engineering et al. 2018. *Data Science for Undergraduates: Opportunities and Options*. <https://doi.org/10.17226/25104>.

Office of Planning and Analysis. 2024. "Common Data Set." Xlsx. <https://docs.google.com/spreadsheets/d/1CfCS76GVbnoWUkERd-mMtJT2EsIN2eVq/edit#gid=360998290>.

Office of the Law Revision Council, United States House of Representatives. 2009. "20 U.S.C. § 1101a: Definitions; Eligibility." United States Code. 2009. <https://uscode.house.gov/view.xhtml?req=1101a&f=tre sort&fq=true&num=12&hl=true&edition=prelim&granuleId=USC-prelim-title20-section1101a>.

Oliver, Jeffrey C., Christine Kollen, Benjamin Hickson, and Fernando Rios. 2019. "Data Science Support at the Academic Library." *Journal of Library Administration* 59 (3): 241-257.  
<https://doi.org/10.1080/01930826.2019.1583015>.

Pon-Barry, Heather, Becky Wai-Ling Packard, and Audrey St. John. 2017. "Expanding Capacity and Promoting Inclusion in Introductory Computer Science: A Focus on near-Peer Mentor Preparation and Code Review." *Computer Science Education* 27 (1): 54-77.  
<https://doi.org/10.1080/08993408.2017.1333270>.

Raj, Adalbert Gerald Soosai, Jignesh M. Patel, Richard Halverson, and Erica Rosenfeld Halverson. 2018. "Role of Live-Coding in Learning Introductory Programming." In *Proceedings of the 18th Koli Calling International Conference on Computing Education Research*, 1-8. Koli Calling '18. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/3279720.3279725>.

Ridsdale, Chantel, James Rothwell, Mike Smit, Hossam Ali-Hassan, Michael Bliemel, Dean Irvine, Daniel Kelley, Stan Matwin, and Brad Wuetherick. 2015. "Strategies and Best Practices for Data Literacy Education: Knowledge Synthesis Report." Dalhousie University.  
<https://dalspace.library.dal.ca/handle/10222/64578>.

Ruhs, Nicholas. 2023. "Developing a Data Fellowship Program and Peer-to-Peer Support Model." *Journal of eScience Librarianship* 12 (1): e625. <https://doi.org/10.7191/jeslib.625>.

Wilson, Greg. 2016. "Software Carpentry: Lessons Learned [version 2; peer review: 3 approved]." *F1000Research* 3 (2016): 62. <https://doi.org/10.12688/f1000research.3-62.v2>.