Computational & Data Science Curriculum Exchange Pilot Program

Linda Akli and Katharine Cahill

Computational Science is a rapidly growing multi- and interdisciplinary field that develops mathematical and computational models and uses advanced computing techniques to simulate these models. Data science deals with vast volumes of data using modern tools and techniques to find unseen patterns, derive meaningful information, and make business decisions. The overarching goal of computational and data science (CDS) is to understand and solve complex problems; and mathematics is central to both. The workforce shortage in many STEM fields is well documented, and it is particularly acute in newly emerging and fast growing domains such as computational and data science.

The National Science Board Policy Companion statement to Science and Engineering Indicators 2018 states "The U.S. can no longer rely on a distinct and relatively small STEM workforce. … The talents of minority groups in the U.S. are perhaps our greatest untapped resource." In 2019, the U.S. population ages 18–64 years old was comprised of Hispanics or Latinos, 18.5%; Blacks or African Americans, 13.0%; Asians, 6.3%; and other racial and ethnic groups combined, 2.8%. Yet, the representation of these groups in science and engineering (S&E) is significantly smaller than their representation in the U.S. population. [1]

CDS Curriculum Implementation Challenges

Computational and Data Sciences (CDS) require the integration of expertise across several disciplines: mathematics, computer science, and domain science and engineering disciplines which poses some challenges to the rapid implementation of CDS curriculum at many institutions. The organization of universities into discipline-oriented departments with budget models tied to those departments is one obstacle. Upper-level courses in each discipline are typically oriented toward majors and often require several prerequisites. Introductory courses in these disciplines often do not include computational modules as part of the curriculum and can be slow to change. New programs may find it difficult to recruit students for their new course offerings given the schedule constraints, prerequisites, and competing well-established course options. Even the existing programs can find it challenging to obtain sufficient enrollment to justify the continued use of limited faculty resources.

Historically Black Colleges and Universities (HBCUs) and High-Hispanic-Enrollment institutions (HHEs) have a major role in advancing African American and Hispanic representation in the STEM workforce. 23.2% of Black or African American graduates who earned a science and engineering doctorate between 2015 and 2019 earned their bachelor's degree from an HBCU and 37.8% of Hispanics or Latinos who received a similar doctorate between 2015 and 2019 received their bachelor's degree from an HHE institution. [1] Thus, it is a priority for these institutions to be able to offer CDS certificates, minors, and majors and why the National Science Foundation funded the Computational and Data Science Curriculum Exchange (C2Exchange) pilot project whose founding institutions include four HBCUs and one HHE.

Despite having strategic plans that include offering CDS curriculum, schools that are teaching intensive, small, primarily undergraduate, and Minority Serving Institutions (MSI) face additional challenges to the implementation of certificate, minor, and major CDS programs.

- Teaching workload is greater than for faculty at research institutions
- Gaps in the expertise necessary to develop and offer all the courses needed for a CDS major, minor, or certificate
- At most, a limited pool of graduate students as teaching assistants
- Less access to local computational and data analytics resources

C^2Exchange Pilot

C^2Exchange is a National Science Foundation (NSF)-funded pilot project to address these curriculum implementation challenges and enable the wide-spread implementation of Computational and Data Science (CDS) curriculum. C^2Exchange integrates core literacy and discipline-appropriate advanced skills in advanced cyberinfrastructure as well as computational and data-driven methods for advancing fundamental research.

The development of the C^2Exchange was informed by data collected via campus visits conducted by the NSF Extreme Science and Engineering Discovery Environment (XSEDE) Education and Broader Engagement programs, outcomes of an Advancing Computational Science at MSIs workshop, requirements articulated by the participating institutions, and the evaluation data collected during the pilot. A competency-based approach guides the development of C^2Exchange courses, and the top-level competencies resulting from this work are identified in figure 1. Each competency area has recommended content and more detailed learning outcomes.[3]

Linda Akli is deputy director of Education Outreach & Training at Sustainable Horizons. Her email address is akli@shinstitute.org.

Katharine Cahill is the education and training specialist at Ohio Supercomputer Center. Her email address is kcahill@osc.edu.

DOI: https://dx.doi.org/10.1090/noti2585

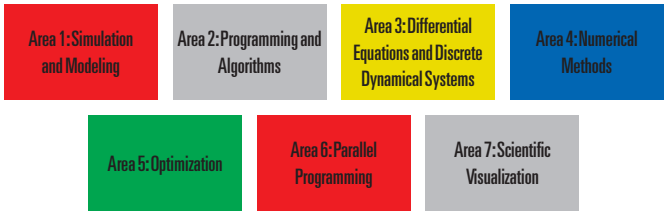

Figure 1. Top Level Computational Science Competencies.

The exchange concept was initially based on the small private online course where students had instructors at their own institutions to guide their engagement with the online material and provide connections to their local program as well as grading in line with their institutional standards. [4] We also focused on the collaborative teaching model where a lead instructor provides support to local instructors as they act as local facilitators for the online course. We have found these models more effective for student retention relative to massively open online courses (MOOCs). [5] This also provides the local instructors with training in the topics to enable them to teach the material or take the collaborative course material and create their own local course.

The C^2Exchange pilot provided the foundation for a collaboratively managed network of five institutions sharing materials and expertise to support their evolving CDS curriculum needs; and it is a model that is expandable to include more institutions and can be replicated by other organizations. Bethune-Cookman (BCU), Clark Atlanta (CAU), Morgan State, and Southern University and A&M College (SUBR) are HBCUs and collectively enroll 183,561 undergraduates with 86% identifying as African American. The University of Puerto Rico at Mayagüez (UPRM), an HHE, has an enrollment of 12,000 undergraduate students with 94% identifying as Hispanic. Students at these institutions will benefit from having access to a broader selection of computational science, which increases their chances of entering the STEM workforce or continuing to graduate school and becoming effective researchers.

BCU, CAU, Morgan, SUBR, and UPRM successfully worked closely together to develop and implement three complete courses:

- Intro to Modeling and Simulation
- Computational Chemistry and Molecular Modeling
- Data Science and Machine Learning (formerly Computational Linear Algebra)

Expertise was shared through management calls, small group meetings, the materials development process, and guest lecture presentations. External evaluation supported continuous improvement as the pilot project evolved and all founding partners were represented in the governance and management of the pilot. All efforts were driven by the needs of the academic institutions. [6]

Exchange Partners' Experiences

CAU is a comprehensive, private, urban, co-educational institution of higher education offering 38 exciting areas of study at bachelor's, master's, specialists, and doctoral levels. One of the strategic goals of CAU is to maintain and develop excellent, innovative academic programs, strategic partnerships, and new modes of teaching and learning to produce graduates who are successful. CAU has a vision of implementing computational and data science (CDS) to undergraduate students in Science, Engineering, and Mathematics Departments in the School of Arts and Sciences. During the pilot stage of the project, the Department of Chemistry participated in the sharing of a "Computational Chemistry and Molecular Modeling" course. Initially, BCU incorporated selected modules of this into their Physical Chemistry course for chemistry majors. Then, BCU used the complete course syllabus and modules for creating a new course that was approved by the university as an elective course for chemistry and biology majors toward building the computational science pathway.

As Maryland's preeminent public urban research university, Morgan State University enrolls a diverse undergraduate population, with over sixty percent supported by Pell Grants and a large percentage of first generation college students. Morgan leveraged an Association of American Colleges and Universities Teaching to Increase Diversity and Equity in STEM (MTIDES) three-year initiative to develop and implement curricula that included cultural competency in teaching and CDS competencies. The C^2Exchange pilot resulted in the complete development of a course, Matrix Methods for Data Science and Machine Learning, that was piloted at the university for two semesters; several modules from the course were piloted by two other members of the C^2Exchange, SUBR and UPRM.

Southern University and A&M College (SUBR) is a member of the Southern University System that has campuses across the state of Louisiana in Baton Rouge, New Orleans, and Shreveport. The Baton Rouge campus (SUBR) offers four-year, masters, doctoral, and professional degree programs. The vision of SUBR is "to provide access and opportunity to students and matriculate graduates who are equipped to excel in a 21st century, knowledge-based, global economy." The C^2Exchange courses provide entry level and advanced training opportunities in computation for students enrolled at SUBR. The "Introduction to Modeling and Simulation" course may be completed by students with minimal mathematics background (prerequisite College Algebra) as a standalone course or in an accelerated "bootcamp" to introduce students to Python or MATLAB for use in a project-based course.

UPRM, the second-largest university campus of the University of Puerto Rico system, is a land-grant, sea-grant, space-grant state university located in the city of Mayagüez, Puerto Rico. Aligned with the current UPRM Strategic Plan, the vision is to implement a Computational and Data

Science Curricular Sequence (CDSCS) for undergraduate students at UPRM majoring in the Sciences from the disciplines of Mathematics, Chemistry, Physics, Biology, and Geology, and all Engineering programs. In the CDSCS, students will take a minimum of 15 credit units to have the curricular sequence as part of their official university transcript. Two courses developed under the previous C2Exchange proposal will be part of the UPRM CDSCS.

Lessons Learned

During the pilot program we identified that there are many avenues for professional support and use of course materials. Exchange materials were used as independent study projects, modules were used as part of existing courses, and whole courses were taught and added to department catalogs. An important finding is that faculty utilize and provide many types of support when they are part of a community of similarly focused instructors. Many times, in small institutions a single faculty member is the only computationally focused instructor in their department or school and has limited opportunities for professional development because they do not have a local network of support. C^2Exchange fills that gap.

The exchange provided insights into both the institutional and pedagogical questions associated with cross-institutional sharing and exchange of curriculum resources. C^2Exchange provides access to curriculum materials aligned with competencies for Computational Data Science education and enhances the capacity of small or under-resourced institutions to offer undergraduate CDS curriculum, minors, and certificates at a low cost and be extensible for future participation by additional resource-constrained institutions.

The C^2Exchange model has the potential to accelerate the implementation of CDS curriculum by sharing existing courses across the participating institutions and preparing a larger number of students to undertake research or enter the workforce where such skills are in short supply. Faculty will be better prepared to teach classes with this content in the future by continuing to participate in the collaborative courses or by integrating the materials into their own curricula. This network strengthens partnerships between institutions, generating more opportunities for collaboration and sharing of resources and effort.

References

[1] National Center for Science and Engineering Statistics, *Women, Minorities, and Persons with Disabilities in Science and Engineering: 2021*, Special Report NSF 21-321, National Science Foundation, Alexandria, VA, 2021, `https://ncses.nsf.gov/wmpd`.

[2] National Academies of Sciences, Engineering, and Medicine, *Data Science for Undergraduates: Opportunities and Options*, The National Academies Press, Washington, DC, 2018, `https://doi.org/10.17226/25104`.

[3] HPC University, *Computational and Data Science Education Competencies*, `http://hpcuniversity.org/educators/competencies`.

[4] K. J. Cahill, S. Lathrop, and S. Gordon, *A Proposed Model for Teaching Advanced Parallel Computing and Related Topics*, The International Conference for High Performance Computing, Networking, Storage and Analysis, 2017.

[5] Steven I. Gordon, James Demmel, Lizanne Destefano, and Lorna Rivera, *Extending Access to HPC Skills Through a Blended Online Course*, XSEDE '15, July 26–30, 2015, St. Louis, MO.

[6] For more information about the courses developed, please see our program website: `https://sura.org/programs/information-technology-2/c2exchange/`.

Linda Akli

Katharine Cahill

Credits

Photo of Linda Akli is courtesy of Linda Akli.
Photo of Katharine Cahill is courtesy of Katharine Cahill.