Building a Computational and Data Science Workforce

Katharine Cahill
Ohio Supercomputer Center
Columbus, OH
kcahill@osc.edu

Linda Akli
Information Technology Initiatives
Southeastern Universities Research Association
Washington, DC
akli@sura.org

Tandabany Dinadayalane
Department of Chemistry
Clark Atlanta University
Atlanta, GA
dtandabany@cau.edu

Ana Gonzalez
Department of Mathematics
University of Puerto Rico Mayagüez Campus
Mayagüez, PR
anacarmen.gonzalez@upr.edu

Raphael D. Isokpehi
College of Science, Engineering and Mathematics
Bethune-Cookman University
Daytona Beach, FL
isokpehir@cookman.edu

Asamoah Nkwanta
Department of Mathematics
Morgan State University
Baltimore, MD
asamoah.nkwanta@morgan.edu

Rachel Vincent-Finley
Department of Mathematics and Physics
Southern University and A&M College
Baton Rouge, LA
rachel_finley@subr.edu

Lorna Rivera
Lorna Rivera Consulting LLC
Lorna@lornarivera.com

Ahlam Tannouri
Department of Mathematics
Morgan State University
Baltimore, MD
ahlam.tannouri@morgan.edu

ABSTRACT

Under-representation of minorities and women in the STEM workforce, especially in computing, is a contributing factor to the Computational and Data Science (CDS) workforce shortage. In 2019, 12 percent of the workforce was African American, while only 7 percent of STEM workers were African American with a bachelor's degree or higher. Hispanic share of the workforce increased to 18 percent by 2019; Hispanics with a bachelor's degree or higher are only 8 percent of the STEM workforce [1].

Although some strides have been made in integrating CDS competencies into the university curriculum, the pace of change has been slow resulting in a critical shortage of sufficiently qualified students at both the baccalaureate and graduate levels. The NSF Working Group on Realizing the Potential of Data Science final report recommends "strengthening curriculum at EPSCoR and Minority Serving Institutions (MSI) so students are prepared and competitive for employment opportunities in industry and academia" [2]. However, the resource constraints and large teaching loads can impede the ability of MSIs and smaller institutions to quickly respond and make the necessary curriculum changes.

Ohio Supercomputer Center (OSC) in collaboration with Bethune Cookman University (B-CU), Clark Atlanta University (CAU), Morgan State University (Morgan), Southeastern Universities Research Association (SURA), Southern University and A&M College (SUBR), and the University of Puerto Rico at Mayagüez (UPRM) are piloting a Computational and Data Science Curriculum Exchange (C^2Exchange) to address the challenges associated with sustained access to computational and data science courses in institutions with high percentage enrollment of students from populations currently under-represented in STEM disciplines.

The goal of the C^2Exchange pilot is to create a network for resource constrained institutions to share CDS courses and increase their capacity to offer CDS minors and certificate programs. Over the past three years we have found that the exchange model facilitates the sharing of curriculum and expertise across institutions for immediate implementation of some courses and long-term capacity building for new Computational and Data Science programs and minors.

KEYWORDS

Computational and data science education, Computational and data science minors, Curriculum consortia, Broadening participation in computational and data science, Computational thinking, Computing education programs

1. CDS WORKFORCE SHORTFALL

There is a well-established need for a STEM workforce with a working knowledge of computational and data science (CDS) [3]. Although some strides have been made in integrating CDS competencies into the university curriculum, the pace of change has been slow, resulting in a critical shortage of sufficiently qualified students at both the baccalaureate and graduate levels. There are significant resource constraints contributing to the slow pace of implementation of undergraduate CDS curriculum that are limiting the production of CDS literate CI-Users from STEM domain sciences. There has been growth in the number of graduate programs in computational science and related data science fields. Though CDS at the undergraduate level would provide a pipeline for graduate programs and prepare students for the workforce, implementation is slower than expected due to several curriculum

DOI: https://doi.org/10.22369/issn.2153-4136/13/1/5

associated challenges. These curriculum challenges include lack of expertise to teach courses, limited access to advanced computing resources to conduct large scale data analytics, and institutional climate for collaborative instruction.

The Ohio Supercomputer Center Education staff track computational and data science (CDS) programs and maintain a list at HPC University [4]. The listing currently shows that there are only twenty-three US institutions with undergraduate computational science programs or minors, seventy-five institutions with graduate level programs, thirty-three institutions with undergraduate data science degree or minor programs, and seventeen with graduate level programs. Only five minority serving institutions are represented in these numbers: Chaminade University, Fisk University, Jackson State University, New Mexico State University, North Carolina A&T University, University of Hawaii at Hilo, and University of San Francisco.

Computational science requires the integration of expertise across several disciplines: mathematics, computer science, and the domain science and engineering disciplines. The organization of universities into discipline-oriented departments with budget models tied to those departments is a major obstacle to the development of interdisciplinary courses in computational science. Upper level courses and even some introductory courses in each discipline are typically oriented toward majors and require a number of prerequisites. Departments are heavily oriented toward providing the courses for their majors, making it difficult to provide courses for non-majors desiring CDS concentrations or minors without committing to a large number of classes. For example, computer science curricula are focused on meeting extraordinary demands for classes in that major, making the education of science majors in the basic computer programming and database management techniques a very low priority.

At smaller institutions and those with smaller STEM programs, departmental workload of faculty reduces their availability for the development of interdisciplinary curriculum. New courses and specialty courses are often not offered due to insufficient number of students to meet minimum enrollment for the course set by the institutions/university. Class exercises require access to advanced computing resources or tools not available locally.

Institutions starting new programs face challenges in recruiting students for the new course offerings given the schedule constraints, pre-requisites, and competing well-established course options. Even the existing programs can be challenged with garnering sufficient numbers of students to justify the continued use of limited faculty resources to offer them with sufficient frequency. These challenges are generally more pronounced at MSIs where budgets are tight, teaching loads are high, faculty with the expertise and experience to teach computational science is either unavailable or insufficient, and the access to appropriate software and hardware environments is limited.

2. CURRICULUM EXCHANGE OVERVIEW

The C^{2}Exchange project is made up of 7 collaborators including four Historically Black Colleges and Universities (HBCU) and one Hispanic Serving Institution (HSI). These partners are collaborating to address the challenges experienced by resource constrained institutions and, in particular, those faced by MSIs.

Computational science expertise at individual institutions is often limited to one or two areas, making it impossible to offer a curriculum aimed at a broad range of STEM majors. The goal of the C^{2}Exchange pilot is to create a network for resource constrained institutions to share CDS courses/curriculum and increase their capacity to offer CDS minors and certificate programs for STEM majors. The proposed exchange model allows the sharing of that expertise across institutions for immediate implementation of some courses and long-term capacity building for the implementation of CDS minors.

The C^{2}Exchange implementation is guided by CDS competencies developed by OSC in collaboration with academic faculty and industry, data collected via campus visits conducted by the XSEDE Education and Broader Engagement programs, outcomes of the SURA-led Advancing Computational Science at MSIs workshop, and the requirements articulated by the participating institutions through a series of surveys, conference calls, and site visits.

3. COLLECTIVE IMPACT

The practical implementation of a computational science program for undergraduates often requires the addition of new courses as well as changes in existing courses so that disruption to the entire curriculum is minimized and the most efficient use is made of faculty and support personnel. The C^{2}Exchange takes advantage of national efforts to integrate CDS into the curriculum and applies distance learning technologies to form a consortium with sufficient critical mass to enable stable CDS offerings at the academic institution partners.

Implementation of computational and data science curriculum is a strategic priority for all of the participating institutions. The C^{2}Exchange builds on the academic partners' prior participation in CDS training and education events and their experience offering hybrid courses. Additionally, the participating institutions completed surveys to identify technology support, experience, current course offerings, and the envisioned institutional benefits of participating in the C^{2}Exchange pilot. Figure 1 presents the goals of the partner academic institutions.

B-CU	• Offer collaborative experiential learning in transdisciplinary data visualization
CAU	• Create Interdisciplinary Computational Chemistry and Biology Certificate
Morgan	• Offering Computational Linear Algebra and Statistics for Data Science Concentration
SUBR	• Updating curriculum w/CDS competencies • Future CDS minor
UPRM	• Updating curriculum w/CDS competencies • Future CDS minor

Figure 1. C^{2}Exchange academic institutions' goals.

The courses each institution receives fills a gap in their current offerings because of one or more of the following scenarios: (a) the course does not exist at their institution; (b) the course exists but no instructor is available to teach the course; (c) there is not sufficient student enrollment to offer the course.

4. C^{2}EXCHANGE FOUNDATION

4.1 CDS Curriculum and Competencies

While developing an interdisciplinary undergraduate minor program in computational science at a group of Ohio institutions, a set of competencies were identified to guide the creation of computational science courses and course materials. The competency-based approach allowed institutions to design their curriculum in a flexible way by integrating portions of the computational science materials into existing courses, creating new courses focused on computational science, or doing a combination of the two.

Maintained by OSC on the HPC University website, the competencies were created by the participating faculty and then reviewed by a business advisory committee that offered some advice on topic emphasis and breadth. Since that time, a number of courses and instructional modules have been developed and tested in a variety of instructional formats. The top-level competencies resulting from this work are shown in Figure 2. Each top-level competency area has recommended content and learning objectives. Figure 3 presents the Area 1: Simulation and Modeling content [5].

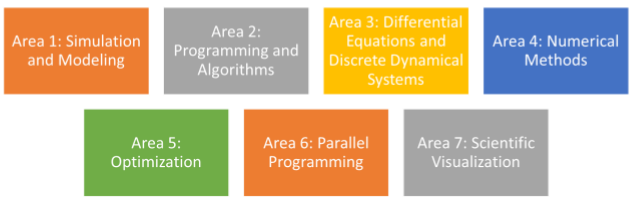

Figure 2. Top level computational science competencies.

Simulation and Modeling

- Explain the role of modeling in science and engineering
- Analyze modeling and simulation in computational science
- Create a conceptual model
- Examine various mathematical representations of functions
- Analyze issues in accuracy and precision
- Understand discrete and difference-based computer models
- Demonstrate computational programming utilizing a higher level language or modeling tool
- Assess computational models
- Verification, Validation, and Accreditation
- Complete a team-based, real-world model project
- Demonstrate technical communication

Figure 3. Computational science competencies — simulation and modeling subareas.

4.2 Blended Education

An online course on parallel computing offered through a partnership between XSEDE and UC Berkeley initially was offered as a Massive Online Open Course (MOOC). XSEDE's evaluation of this initial offering was consistent with the research indicating that MOOCs were ineffective and resulted in very high dropout rates and poor learning outcomes [6]. XSEDE's course was subsequently offered as a Small Private Online Course (SPOC) with collaborating faculty from a variety of higher education institutions that offered the course for credit.

This subsequent offering using a blended learning approach was found to be an effective educational model for teaching computational science courses at a variety of institutions. The lecture materials are delivered as online videos followed by quizzes to gauge student comprehension. Computing exercises are completed by all students to apply the knowledge gained in the lectures. Students may also complete a final project where they use the parallel computing techniques reviewed in the class to an application of their own choice. Local faculty members serve as advisors to these students and were responsible for grading as well as altering exercises as they saw fit for their students.

The advantages of this model are the course materials are readily available and the local instructors can offer the course with a minimum of preparation time. Students gain by being offered a course that was not previously available at their institution. The face-to-face component enabled students with a weaker computing background to gain the help they needed to complete the assignments. Moreover, keeping the classroom environment gives the students the focus they needed to complete the course. This offering had 91% completion rate compared to the MOOC with only 5% completion rate.

5. COURSE DEVELOPMENT AND DELIVERY

In the C^{2}Exchange pilot project, the participating institutions agreed to offer one course and participate as a recipient of at least one of the courses offered by the other consortium members. Each institution will take the lead in creating the first version of each course and agree to have their faculty be the local instructor for courses led by other institutions. Each institution will work with their administration to institute a formal minor or certificate program in computational science that uses both their local courses and the shared, online courses as part of the curriculum.

The courses offered during the pilot are drawn from the undergraduate competencies identified in Figure 2 and available at HPC University [5]. Initially only four courses will be offered to provide enough time for adapting courses to be delivered as a Small Private Online Course and to offer each course more than once for an effective evaluation.

Introduction to Modeling & Simulation is a 4-credit course that introduces the principles of modeling and simulation combined with an introduction to programming principles and skills using Python. It covers the construction, development, and study of mathematical representations of different classes of models; basic examples; techniques for fitting a function to an experimental data set; and selected case studies are included. This course may be offered to students well into their STEM major who are interested in developing computational skills.

The primary goal of the course is to introduce basic concepts of computational science to a diverse student body. The aim is not to produce experts in computational science but to provide the tools and skills that can benefit the personal and professional lives of individual and allow then to better collaborate with computational scientists. The course is targeted to students of all majors.

Computational Linear Algebra topics include Review of MATLAB and Basic concepts from Linear Algebra I; Cholesky Decomposition; Singular Value Decomposition; Principal Component Analysis; Matrix Approximation; Maximum Likelihood (Linear and Logistic regression); Support Vector Machines; Clustering algorithms, Gaussian Mixture Models; Dimensionality reduction techniques; and as time permits, applications. Most STEM majors are required to take a linear algebra course for their degree programs. Infusing more computation into their background through topics involving computational linear algebra will broaden their training and help prepare them for graduate and professional schools as well as the workforce.

The course was renamed Matrix Methods for Data Science and Machine Learning and will be offered for Mathematics, Computer Science, Physics, and Engineering majors at Morgan State University; this will result in students being better prepared to confront emerging challenges of a new data-driven world and to respond to the workforce needs; the course will open doors for graduate work in mathematical and computational sciences graduate programs and job opportunity in Data Science, Machine Learning, and AI fields.

Introduction to Computational Chemistry and Molecular Modeling is comprised of lectures and labs to introduce the concepts of computational chemistry and molecular modeling and their applications in chemistry and biochemistry. This course is mainly for upper level undergraduate students of chemistry and biology majors, but students in any STEM discipline can take this course. The students in this course will learn how to use computers in chemistry as well as related fields like molecular sciences, drug design, biomedical, and materials science.

The Transdisciplinary Data Visualization learning experiences are designed to provide students with the procedures and principles for the design and deployment of interactive visual representations of large professionally collected datasets of societal relevance. The pre-requisite is a senior academic status with prior statistics or computing course at 200 level or above. The course is to be optimized for 4-week, 8-week and normal semester durations as well as face-to-face and online learning environments.

The courses reflect a blend of core competencies for an undergraduate computational science certificate program as well as the expertise of our partners and are based on currently offered courses to reduce development time. Introduction to Modeling and Simulation, Data Visualization, and Computational Linear Algebra are foundational courses for any computational science program, while Computational Chemistry and Molecular Modeling is a discipline specific course.

Through course development discussions, the syllabi and materials for the courses are reviewed collectively prior to implementation. Thus, all participating institutions are providing input on content and identifying how the material will fit into their curriculum.

Table 1. Courses received each year.

	Fall 2019	Fall 2020	Fall 2021
Modeling & Simulation	x	x	x
Matrix Methods Machine Learning		x	x
Computational Chemistry	x	x	x
Data Visualization (Winter term)			x

6. GOVERNANCE — MANAGEMENT AND OWNERSHIP

Providers such as XSEDE, Virtual School of Computational Science and Engineering, and the Great Lakes Consortium Virtual School successfully offer multi-site training and blended online courses in Parallel Programming and other High Performance Computing topics. However, their course offerings don't fully support the goals of C^{2}Exchange.

1. They offer a just a few courses and don't cover all the topics needed to develop a CDS minor for STEM majors.
2. The sites or receiving institutions do not influence what or when courses are offered, thus making it difficult for local or receiving sites to plan out their curriculum offerings.
3. There isn't knowledge exchange between the providers and the participating sites, thus lacking intentional capacity building at the receiving sites.

The C^{2}Exchange is designed to create a scalable network of institutions that can collectively offer CDS minors, concentrations, or certificates with minimal investment. The academic institution partners collaboratively create and maintain a multi-year plan identifying what courses will be offered, when the courses will be offered, and who will be the lead for each course. The active management role by the academic institution partners provides the potential for stable CDS curriculum offerings at their institutions and a pathway to the implementation of minor and certificate programs.

Without receiving credit, it will be difficult to recruit students to enroll in the course, and there would be little incentive to complete the course. Each of the participating institutions has mechanisms for setting up courses for credit as long as there is a faculty member responsible at the receiving institutions. How the courses are listed in the catalog vary depending on whether there is an existing course or there are options to offer it as a seminar or special topics course during the initial pilot. To date, no impediments at the lead or other institutions offering the courses have been encountered. The development of a C^{2}Exchange Memorandum of Agreement or similar agreement is being explored to ensure courses continue to be offered.

7. EVALUATION

C^{2}Exchange is employing a robust evaluation designed to provide formative information to guide program improvement as well as a summative assessment of program effectiveness and impact. The ultimate goal of the evaluation is to validate and document the effectiveness of the model exchange for enabling CDS minor and certificate programs and disseminate findings through publication and presentation. The evaluation will utilize a Values-Engaged Educative Approach (VEE) [9]. The VEE approach, developed with NSF-EHR support, defines high quality STEM educational programming as that which effectively incorporates cutting edge scientific content, strong instructional pedagogy, and sensitivity to diversity and equity issues. In the VEE approach, a key role of the evaluator is to work closely with program implementers to promote their understanding of program theory, implementation and impact.

The evaluation is designed to answer four questions:

1. Implementation: Is the C^{2}Exchange project being implemented on schedule and as planned?
2. Effectiveness: Are key components of the C^{2}Exchange model (e.g. enrollment, retention, curriculum development, course exchange, Annual Project Meeting, etc.) operating effectively and for whom? How might they be improved?
3. Impact: What outcomes (e.g. scientific impact, gains in scientific knowledge, improved technical skills, certification, student employment outcomes, increased institutional capacity, etc.) are associated with participation in the C^{2}Exchange programs? How does impact vary across groups? What is the value-added of participation in the C^{2}Exchange program?
4. Institutionalization: How and to what extent are elements of the C^{2}Exchange programs becoming institutionalized to ensure sustainability of program components? What opportunities and barriers exist?

Two rounds of evaluation have been conducted with the exchange partners to cover the fall semester offerings of 2019 and 2020. The survey found that faculty view C^{2}Exchange participation as a valuable professional development activity with added benefits for local students and respondents generally report higher levels of experience in advanced computing areas compared to 2019.

A survey respondent commented, “The C^{2}Exchange has provided excellent professional learning experience for me. The Exchange Model is being expanded at my institution in a way to promote knowledge sharing, trust, reciprocity, and collaboration.”

We hope that with expansion of exchange we are able to measure student impact directly in our future work.

8. CONCLUSION

The pilot so far has demonstrated that the C^{2}Exchange facilitates the exchange of strengths in the computational and data science curriculum available at the partnering institutions. Requirements for governance and inter-institutional agreements are being explored, and the initial implementation structures will be determined by the end of the pilot in summer 2022.

9. ACKNOWLEDGMENTS

This work is funded by the National Science Foundation Grant Cybertraining:CIU: Computational and data science literacy exchange (CSE-1829717).

10. REFERENCES

[1] National Science Board, National Science Foundation. 2021. The STEM Labor Force of Today: Scientists, Engineers and Skilled Technical Workers. Science and Engineering Indicators 2022. NSB-2021-2. Alexandria, VA. Retrieved from https://ncses.nsf.gov/pubs/nsb20212

[2] 2016. Realizing the Potential of Data Science: Final Report from the National Science Foundation Computer and Information Science and Engineering Advisory Committee Data Science Working Group. (Dec. 2016).

[3] National Academies of Sciences, Engineering, Medicine. 2016. Future Directions for NSF Advanced Computing Infrastructure to Support U.S. Science and Engineering in 2017-2020. Retrieved from http://sites.nationalacademies.org/CSTB/CompletedProjects/CSTB_087924

[4] HPC University: Graduate and Undergraduate Programs. Retrieved from http://hpcuniversity.org/students/undergraduatePrograms/

[5] HPC University: Competencies. Retrieved from http://hpcuniversity.org/educators/competencies/

[6] Gordon, S. I., Carey, K. and Vakalis, L. 2008. A Shared, Interinstitutional Undergraduate Minor Program in Computational Science. *Computing in Science and Engineering* 10, 5 (Sept./Oct. 2008), 12–16, DOI: https://doi.org/10.1109/MCSE.2008.127

[7] Gordon, S. I., Demmel, J., Destefano, L. and Rivera, L. 2016. Implementing a Collaborative Online Course to Extend Access to HPC Skills. *Computing in Science & Engineering* 18, 1 (Jan./Feb. 2016). DOI: https://doi.ieeecomputersociety.org/10.1109/MCSE.2016.6

[8] Hughes Darden, C., Akli, L., Bapna, S., Bhattacharya, P., Emdad, A., Hargett, S., Ellington, R., Nkwanta, A. and Zaveri, J. (2019). Interventions Addressing Recruitment and Retention of Underrepresented Minority Groups in Undergraduate STEM Disciplines. *Culturally Responsive Strategies for Reforming STEM Higher Education: Turning the TIDES on Inequity*. (Jan. 19, 2019). DOI: https://doi.org/10.1108/978-1-78743-405-920191014

[9] Greene, J., DeStefano, L., Burgon, H. and Hall, J. 2006. An Educative, Values-engaged Approach to Evaluating STEM Educational Programs. *New Directions for Evaluation* 109 (Mar. 2006), 53–72. DOI: https://doi.org/10.1002/ev.178