Computational and Data Science

Executive Summary

We propose a campus research theme in Computational and Data Science. This research field has emerged as an important companion to theoretical and experimental research in natural sciences, engineering, and social sciences. Due to recent advances in high-throughput measurement technologies in nearly every active research area, as well as increased computational hardware capabilities, we now have unprecedented opportunities and access to terabyte, petabyte, and even exabyte datasets. Computational and Data Science researchers are addressing the need for developing next-generation computational models and methods to analyze data which, in turn, lead to predictive theories and more refined experimental investigations.

Definition of Thematic Area

Computational and data science is the study of mathematical and statistical algorithms, prediction techniques, modeling methodologies, data collection, data analysis, and visualization to address the massive amounts of data now being generated through experiments and observations across all areas of science and engineering. Computational and data science fits squarely within the theme identified by the SAF Working Group entitled, “Information, Computational, and Data Sciences, and Engineering.” This proposed theme is a conglomeration of the following SAF initiatives (listed alphabetically).

- Applied Mathematics
- CHASE
- Chemistry & Chemical Biology
- CIDER
- Cognitive Science
- Electrical Engineering & Computer Science
- Hard Rock Institute
- Materials Science & Engineering
- Mechanical Engineering
- MIST
- Next Generation Materials
- Political Science
- Center for Statistical and Quantitative Research
- School of Management & Economics
- SpARC
- UC Merced Center for Theory & Computation
- UC Merced’s Library

Intellectual components

The research field of Computation and Data Science is inherently interdisciplinary and has become a crucial companion to theoretical and experimental research in natural sciences, engineering, and social sciences. Computational and Data Science researchers are addressing the need for developing next-generation computational models and methods to analyze data which, in turn, lead to predictive theories and more refined experimental investigations. As indicated by both federal and private research funding agencies nationwide, this theme of “data-enabled science and engineering” is now a thriving area of interdisciplinary research all its own.

For example, the National Science Foundation's new Research Traineeship Program has "Data-Enabled Science and Engineering" as its only priority research theme.
Achievements in this research field vary widely because of computational and data sciences' role across different research domains. For example, computational fluid dynamics is a well-established field and it is considered an integral part of fluid dynamics research along with theoretical and experimental approaches. The same is true for areas such as electromagnetics, chemical mass transport, molecular dynamics, and other areas of science and engineering. Similarly, computational and data science have become crucial components of research programs in areas such as bioinformatics and cognitive science as well. One indication of the emerging importance of computational and data sciences in cognitive science is President Obama's BRAIN Initiative.

UC Merced's role

This theme reaches across all three schools as demonstrated by the faculty participants in this proposal. Through establishing a Computational and Data Science research theme, we will capitalize on the individual impacts that various programs are making, and establish a cohesive, campus-wide signature of excellence that will gain national and international attention. At present, there are many disparate computational training efforts taking place across the campus. By consolidating these individual efforts under one research theme, we will simultaneously:

1. Channel administrative and faculty resources in a strategic manner that will encourage synergies in research and teaching,

2. Avoid unnecessary redundancies (i.e. "reinventing the wheel") in graduate training pedagogy in computational and data science,

3. Lead to efficient sharing of research computing infrastructure resources and instructional space.

It is noteworthy to mention that this research theme is poised to flourish and grow right now. Extensive capital planning is not needed. Capital planning that is needed can be shared across several research and teaching units. In fact, establishing this theme provides a mechanism to make more efficient use of current and future resources. Additionally, because Computational and Data Science complements theoretical and experimental inquiry and knowledge discovery, a cohesive group of faculty in this area is crucial for promoting and developing a comprehensive research profile in all areas of natural sciences, engineering, and social sciences on this campus.

Faculty participation

The following units and graduate groups might participate in this research theme.

- Applied Mathematics (unit + graduate group)
- Chemistry & Chemical Biology (unit + graduate group)
- Cognitive and Information Sciences (unit + graduate group)
- Electrical Engineering & Computer Science
- Environmental Systems
- Materials Science & Engineering

- Mechanical Engineering
- Quantitative Systems Biology
- Physics (unit + graduate group)
- Political Science (unit + graduate group)
- Psychological Sciences (unit + graduate group)

Special Programmatic Needs

For research in computational and data sciences, it is absolutely crucial to have and support research computing resources on campus. Research computing resources include hardware and software for a few small- and medium-scale computing platforms for development and testing of code, training of students, and performing scale studies required for transitioning to large-scale computing platforms. For large-scale computing needs, faculty will make use of national supercomputing centers such as XSEDE (https://www.xsede.org/). In addition, research computing administration support staff will be needed to provide maintenance to the hardware and software, ensure stability in data storage, and monitor the security of the campus research computing systems.

At present, research computing resources are lacking in coordination and organization leading to unnecessary expenses incurred and ineffective use. Most research computing on campus is done with individual faculty vying for space, power, and cooling resources within the limitations of campus resources. This research theme will provide an identity on this campus to develop innovative initiatives for sharing these required research computing resources across all of the computational and data sciences faculty. This sharing strategy may be highly effective for purchasing software licenses, for example. The participating faculty will be able to provide the coordination and re-organization needed to make most effective use of the resources available. Moreover, the participating faculty will provide a means to prioritize research computing needs for the campus and help to develop realistic plans for hardware and software upgrades as well as policies and procedures for sharing research computing resources.

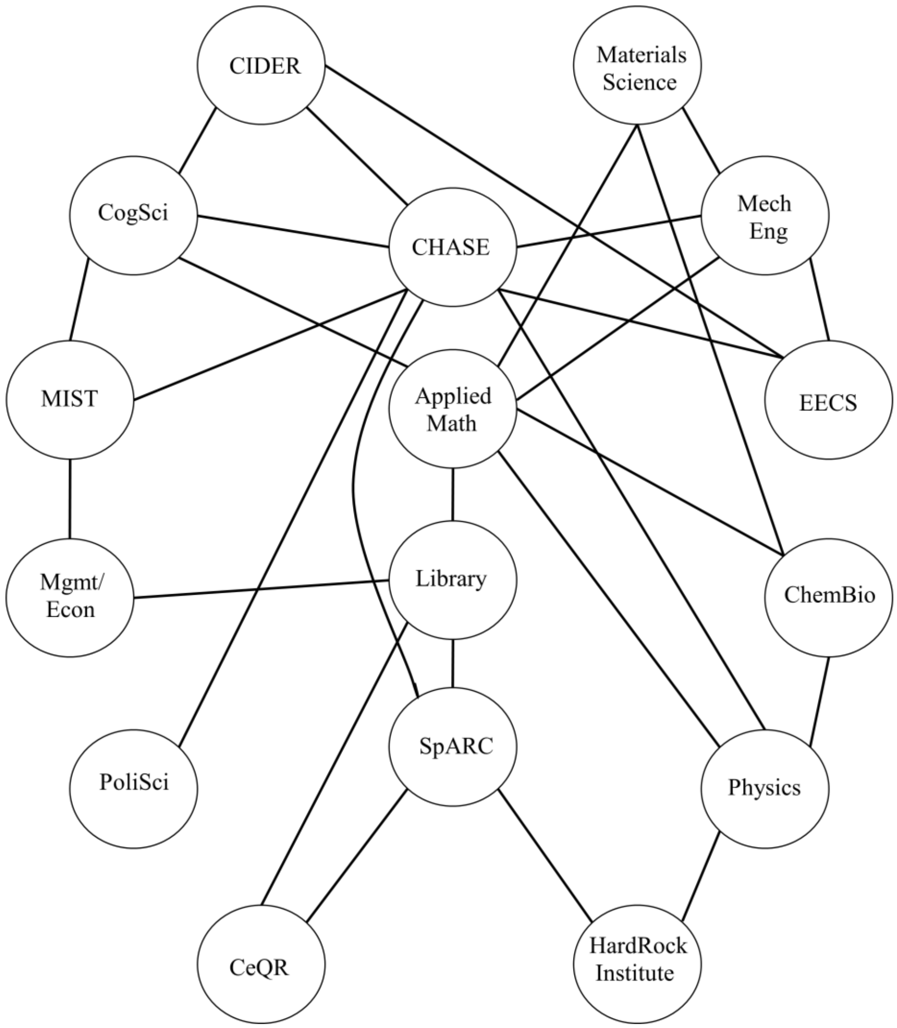

Figure 1. A network diagram illustrating some of the existing and potential collaborative research and teaching connections across units and programs in the Computational and Data Science Research Theme.