CaRDS activities in 2020

Our main business this year was promoting and completing our plans for the ETSU M.S. degree in Applied Data Science (1). Also, CaRDS contributed to finding new avenues for research computing at ETSU (2). And, while there was nothing resembling a regular CaRDS seminar this year, we had nevertheless one seminar talk (3).

1) Steps towards the ETSU Master's degree in Applied Data Science

The central data-related project pursued by CaRDS this year was getting our plans for a Master's degree in Applied Data Science (M.S.A.D.S.) on their way to the Tennessee Higher Education Commission (THEC). Plotting the respective activities versus the twelve months of this year would yield a U-shaped curve. We made good headway in at the beginning of the year, were intermittently slowed down by the health crisis and the associated financial crisis in the A&S college, and resumed our previous pace in the fall. The Letter of Notification (LON) for THEC was submitted in mid-October.

Here is a summary of this year's main stations on the way to the M.S.A.D.S.:

On January 14, CaRDS met with Jason Carter, Director of Analytical Services, Integrated Solutions Health Network, Ballad Health, and Dr. Matthew Loos, Chief Academic Officer, Ballad Health, to discuss Ballad's potential role in establishing and sustaining the M.S.A.D.S. The conversation centered on real-world projects to be provided by Ballad for students of Applied Data Science, hosting data science interns, and designing a health sciences concentration within the M.S.A.D.S. A summary of the discussion is found in the minutes of the meeting sent on January 15. Later interactions with Ballad revolved mostly around compliance issues associated with our students' use of Ballad data.

On February 4, CaRDS board members attended a meeting on the M.S.A.D.S. among the deans of the participating colleges (A&S, CBAT, Public Health), or their representatives, and other stakeholders in the program. This meeting was hosted by Dr. Sharon McGee. The plans presented by CaRDS met with approval. The discussion centered mostly on the question of course delivery (on-ground versus on-line), as well as various items of the proposed budget.

In April Dr. Bishop called a meeting with educational consultants of the company EAB Global Inc. EAB which I attended. EAB was tasked by ETSU to assess graduate programs existing at our campus, and also possible future additions to our graduate portfolio. I requested that EAB includes data science in their evaluation and received the result a few months later. EAB made a strong recommendation for installing an M.S.D.S. degree at ETSU. EAB's diagnostic ranked Data Science among the programs of maximum promise, characterized by a strong labor market with a simultaneous strong competitive opportunity. This assessment has entered the LON.

On August 6, I contacted the new interim Dean of the A&S college, Dr. Joe Bidwell, and inquired about our possibilities to continue the M.S.A.D.S. project. His answer was spontaneous and encouraging. On August 19, CaRDS representatives met with Dr. Bidwell. The focus of the meeting was on the program budget. Dr. Bidwell was strongly in favor of renewing the commitment of the A&S college to the proposed new data-science faculty position in the math department, as supported by Dr. Anderson prior

to the summer of this year. Dr. Bidwell approved the project after having consulted with the deans of the other two participating colleges, and also with President Noland.

After inspection by the office of Dr. Bill Flora, the LON underwent a final revision and was submitted on October 15. So far it has passed through the following stages: ETSU Library, Department of Mathematics and Statistics, A&S College, School of Graduate Studies, ETSU Academic Council, ETSU University Council. It is scheduled to be presented to the Board of Trustees in February 2021.

2) Research computing at ETSU

By the end of July, the partnership between ETSU and the Advanced Computing Facilities (ACF) at Oak Ridge National Laboratories (ORNL)/University of Tennessee at Knoxville (UTK) came to an end. David Currie led an ITS initiative to find a successor agency. David and I had conversations with representatives of four organizations:

- *Joint Institute for Computational Sciences (JICS), UTK -*
  JICS is a joint venture between UTK and ORNL. According to Victor Hazlewood, Chief Operating Officer at JICS, ETSU researchers will have an option to buy JICS computer nodes at $10,000 per node (2 x 24 cores) for the duration of three years. Additional costs at about $2,000/year will comprise maintenance and access to support by the OIT at UTK. The machines available to ETSU will be those previously administered by ACF (now ISAAC, see https://oit.utk.edu/hpsc/).

- *Amazon Web Services (AWS)* –
  AWS provides on-demand cloud computing platforms to individuals, companies, and administrative units. In particular, the Amazon Elastic Compute Cloud (EC2) that allows users to run their own computer applications on virtual computers, is open to ETSU researchers on a pay-as-you-go basis. ETSU currently benefits from the AWS Educate program which offers student research credits and also supports teaching faculty. A modest amount of computing funds is free of charge.

  Customers can choose between different *instance types* (Amazon EC2 Instance Types - Amazon Web Services), i.e. they can design their virtual machine of choice in terms of memory, storage, run time, and number of nodes. Research computing as practiced at ETSU, e.g. in the areas of computational chemistry, materials science, or biological modeling, is a costly venture on the AWS platform. To make use of the AWS offer, researchers may try to secure grants awarded by AWS. Also, the ETSU Research Development Committee (RDC) has supported AWS-based computing.

- *Google Cloud Platform (GCP) -*
  GCP is a suite of cloud computing services that runs on the Google infrastructure (https://cloud.google.com/). In all respects that matter for the ETSU computing venture, GCP seems to be closely comparable to AWS. Off-hand, the prices for on-demand computing are higher, but discounts are available.

- *eXtreme Science and Engineering Discovery Environment (XSEDE) -*
  XSEDE is an international network that encompasses shared computing resources, data, and related technical expertise. Recently, ETSU has joined the XSEDE Campus Champion program

(https://www.xsede.org/community-engagement/campus-champions), and a request for a pilot allocation has been submitted to the Pittsburgh Supercomputing Center (PSC).

Once granted, the pilot access can be used by interested ETSU faculty to run some benchmark jobs at the PSC. The respective results may be used to request an individualized start-up allocation (up to 200 000 core hours). The next (highly competitive) step is submitting a full research allocation request, using the results generated in the start-up stage. Researchers can submit a full research request once every year. Allocation policies are summarized here: https://portal.xsede.org/allocations/policies. The XSEDE Campus Champion program may offer, at this time, the most realistic option to continue the computational research at ETSU.

3) CaRDS Seminar

We hosted one seminar talk this year:
Dr. Bjarne Berg, Professor of Computing Sciences and Professor of Mathematics at Lenoir-Rhyne University, spoke about *Advanced Analytics implementations in the Financial World* on February 26 2020.