AC 2007-1179: DEVELOPING A MULTIDISCIPLINARY ONLINE CYBERINFRASTRUCTURE COURSE THROUGH PROJECT-CENTRIC BIOINFORMATICS

Lionel Craddock, Bluefield State College

Daphne Rainey, Virginia Bioinformatics Institute

Susan Faulkner, Virginia Bioinformatics Institute

Frank Hart, Bluefield State College

Martha Eborall, Bluefield State College

Lewis Foster, Bluefield State College

Stephen Cammer, Virginia Bioinformatics Institute

Betsy Tretola, Virginia Tech

Bruno Sobral, Virginia Bioinformatics Institute

Oswald Crasta, Virginia Bioinformatics Institute

Bruce Mutter, Bluefield State College

Developing a Multi-disciplinary Online Cyberinfrastructure Course through Project-Centric Bioinformatics

Abstract

It is increasingly difficult for teaching to keep pace with rapid advances in technology, especially at the interface of several disciplines. We describe here the development and implementation of an interdisciplinary bioinformatics course focused on preparing the future scientific workforce. Central to the course is a project-centric teaching paradigm to engage students in applying the concepts of cyberinfrastructure through the integration of the disciplines of biology, computer science, mathematics, and statistics in the field of bioinformatics. In this project, Bluefield State College (BSC) professors and their students were introduced to the concepts of cyberinfrastructure (CI) through the application of genomics software tools and data. The cornerstone of the project-centric approach was the development and implementation of educational modules centered on applying a transdisciplinary approach to specific and typical challenges that are faced by current scientists in the area of pathosystems biology (host-pathogen-environment interactions). The course modules were further modified by BSC to fit their students and training objectives. We report here the first implementation of the CI course and a summary of our initial observations to aid others in implementing similar courses. Specifically, we discuss materials developed at Bluefield and implementation of the Center for Applied Research and Technology (CART) Course Management Service (CMS) at BSC in the delivery of the course as well as the assessment.

Introduction

It is increasingly difficult for teaching to keep pace with rapid advances in science and technology, especially at the interface of several disciplines. The rapid and continued developments in information technology are now the driving forces of many of these advances. To address this challenge, the National Science Foundation's (NSF) Assistant Director for Computer and Information Science and Engineering (CISE) convened a Blue Ribbon Panel to explore the trends in information technology and to make strategic recommendations on programs that NSF should award in response to advances in information technology. The summary report from this event, now referred to as the Atkins Report (Atkins et al., 2003) on Cyberinfrastructure (CI), launched an NSF funded program towards the integration of Information technology (IT)-enabled systems, tools, and services to create a national cyberinfrastructure directorate. The premise was that integration of the multitude of tools and services into a national cyberinfrastructure directorate would enable access to multidisciplinary information for many individuals and groups that had previously been marginalized and thereby revolutionize the way science is done. Specifically, the goals were to harness the full power of cyberinfrastructure for discovery, learning, and innovation across and within all areas of science and engineering in the preparation of a workforce with the knowledge and requisite skills needed to create, advance, and exploit cyberinfrastructure over the long-term. Thus a Cyberinfrastructure Training Education Advancement and Mentoring (CI-TEAM) program was created.

CI-TEAM Demonstration Project

In 2005, the Virginia Bioinformatics Institute (VBI) at Virginia Tech University was awarded a CI-TEAM Demonstration project. The CI-TEAM members of this project consist of Bluefield State College (BSC), Bluefield, West Virginia, Galileo Magnet High School (GMHS), Danville Virginia, and VBI. The CI-TEAM Demonstration project began in January 2006. The initial plan for the two-year project was to develop material to create Cyberinfrastructure courses that would be implemented and evaluated by both BSC and GMHS in 2007. The VBI role was to develop course modules based on ongoing projects utilizing cutting edge bioinformatics tools and genomics results to allow for the introduction of the concepts of cyberinfrastructure to students and faculty of each institution to the concepts of cyberinfrastructure. BSC professors and GMHS teachers were charged with the task of developing and bringing together materials to supplement the modules and tailor the information to the students at their respective institutions. BSC prepares many non-traditional students for challenging careers, graduate study, informed citizenship, community involvement, and public service in an evolving global society. The college offers undergraduate liberal arts and professional programs in applied sciences, business, education, humanities, social sciences, engineering technologies, and allied health sciences. This paper describes the work that was done to develop, implement, and assess the first offering of a (CI) course at BSC.

Central to the course is a project-centric teaching paradigm to engage students in applying the concepts of cyberinfrastructure by integrating the disciplines of biology, computer science, mathematics, and statistics through bioinformatics. An important goal was to demonstrate the connections between these often-disparate fields.

We report here the implementation of the first CI course and a summary of our initial observations to aid others in implementation of similar courses. Specifically, we discuss some of the materials that were developed, the use of the Center for Applied Research and Technology, Inc. (CART) Course Management Service (CMS) for online course delivery and some of the pedagogical considerations important to course implementation.

Teaching/Learning Partners

Virginia Bioinformatics Institute (VBI), a Commonwealth of Virginia shared resource established at Virginia Tech in July 2000, serves as a flagship bioinformatics research institute wedding cutting-edge biological research with state-of-the-art computer science. Used synergistically in a diversified research portfolio, these tools catalyze new knowledge and economic development for the Commonwealth of Virginia and beyond. The Institute's esteemed faculty and staff encourage research collaboration to increase the understanding of molecular, cellular, and environmental interactions that affect human health, agricultural systems, and the environment.

Bluefield State College (BSC), located in Bluefield, West Virginia, prepares many non-traditional students for challenging careers, graduate study, informed citizenship, community involvement, and public service in an evolving global society. The college offers undergraduate liberal arts and professional programs in applied sciences, business, education, humanities, social sciences, engineering technologies, and allied health sciences.

CART at BSC is a vehicle for entrepreneurial success. It uses its Applied Research Assistant Program (ARAP) to provide teams of engineering technology students for our School of Engineering Technology and Computer Science (SETCS) with in-house internship experience and the School with a source of increased funding through other CART research contract operations, the operation of the CMS is a fee-based self-sustaining business operation; as are the ongoing applied research projects for industry and agencies; and partnerships with other colleges, universities, industry, and government.

Teaching/Learning Paradigm

As VBI has proven through its foundation and subsequent research, the need for a new workforce mixing science and technology professionals is growing and has been recognized (Atkins et al., 2003). This new workforce will be comprised of individuals with expertise in a particular science domain, as well as in computer science and mathematics. The faculty and partners at VBI are already aware of the need for interdisciplinary training of graduate students. Interdisciplinary training allows students to be more productive in research projects early in their graduate training and effectively prepares them for more productive and satisfying careers in our new knowledge-based economy.

A typical research project requires multiple disciplines coming together to work on a specific research problem. This demonstration project at the undergraduate level brought professors together from different disciplines to utilize this new teaching/learning paradigm, develop project-centric coursework, deploy the project-centric course to BSC students, and form the basis for continuous improvement in the curriculum as needed to reflect this new teaching/learning paradigm. The BSC professors were provided the necessary mentoring and tools to develop and deploy a course that facilitates multidisciplinary teaching/learning, thus advancing the preparation of current and future generations of scientists, engineers, and educators.

Project-Centric Cyberinfrastructure Education: Course Development

Implementation of a project-centric teaching paradigm was aimed at engaging students in applying the concepts of cyberinfrastructure. During the process of course development and delivery, we made use of the rapidly increasing volume of biological data and accompanying bioinformatics tools and they served as valuable teaching resources.
For this course, real scenarios were used to design projects such that the solutions required contributions from personnel with diverse areas of expertise such as molecular biology, microbiology, and bioinformatics. Accessing databases and analysis tools often requires minimal specialized computer skills (i.e. accessing the Internet); the challenge becomes helping students understand the context of the biological problems that can be addressed with these tools (Greene and Donovan, 2005). Accordingly, we chose to focus the course on learning modules centered on several key pathogens of interest to bio-defense and to public health. The objective was to stimulate interaction and participation and ingrain the CI concept through role playing activities and presentations. Roles were developed around professionals that would come into play should an outbreak occur, such as Center for Disease Control specialists, researchers involved in vaccine and/or drug development, hospital physicians, microbiologists, and evolutionary

biologists). Students work in different roles, accessing various data to come to a joint approach for addressing and solving the problem. A focus was placed on the importance of forming transdisciplinary teams and bioinformatics data and tool use. While background was important to the depth of understanding a problem, it was not intended for students to require extensive expertise in any one area.

Project-Centric Teaching using Learning Modules

The key concept of this project is to provide an understanding of CI to students through a problem-based approach rather than a discipline-centric view. The main objective was to stimulate interaction and participation and instill the CI concept through role playing activities and presentations (Rainey et al., 2007). At BSC, the first cohort of students participated in three learning modules to obtain a sense of integrating biological data through use of CI in realistic scenarios. These scenarios involved known human and animal pathogens in current VBI research. The data were incorporated into three modules focused on analyzing a novel strain of *Bacillus anthracis*, the causative agent of anthrax; identifying novel drug targets for *Rickettsia* species, the tick-borne parasites responsible for typhus, Rocky Mountain Spotted Fever and Lyme's disease; and identifying vaccine candidates against *Brucella abortus*, causative agent for *brucellosis* in cattle.

Bioinformatics Tool Usage

During the course the core bioinformatics tools were introduced, focusing on sequence searching and retrieval from National Center for Bioinformatics (NCBI), sequence comparison, sequence manipulation in the form of multiple sequence alignments, phylogenetic tree production, and genome alignment software in Toolbus (He et al., 2005). A variety of software sources were introduced for sequence alignment and phylogenetic tree production and viewing, however BSC professors tested and chose an appropriate source for its own use.

Specific Example of Project Activity

The course was constructed around one live meeting per week, and 4 hours of online activity through the CART CMS required assignment submissions, discussion forums, software interface, quizzes and readings. One of the proposed projects was based on a bio-defense scenario requiring an emergency response to a pathogen outbreak projected to take place some time in the future. The project required that students draw from Science, Technology, Engineering, and Mathematics (STEM) knowledge domains to provide evidence for their organism classification. A team-oriented approach was necessary to address the project challenge, thereby reinforcing CI-TEAM integration principles. Web-accessible tools and data sets were employed to enable students to become familiar with the use of distributed data sources. Project reports were generated and presented by each project team, thus requiring students to practice effective communications skills in a collaborative environment. The use of CART CMS provided for effective communication among students, teachers, and VBI researchers. VBI faculty made visits to the classroom to talk with students about tool use, course modules, and CI perspective in information technology and careers. BSC faculty maintained close contact with VBI researchers throughout the course. Near the end of the course, BSC students made a visit to the VBI facilities for a tour and discussion of course progress. Thus, the project required team collaboration and consultation between and BSC faculty and VBI researchers. For delivery of the learning modules BSC emphasized: (1) introduction to the scenario, (2) introduction to the multiple disciplines, (3) development of strategies, and (4) implementations and conclusions. The first two steps were conducted in the classroom and online, while the last two were conducted jointly between BSC and VBI. At VBI the CI- TEAM provided hands-on workshops for BSC faculty and allowed for student discussions with experts in the respective disciplines.

During the first part of the semester, BSC professors laid the foundation for working with the learning modules. Background material needed to understand the pathogens, as well as the tools to be used for analysis, were presented.

One example of the background material that was presented during the course was an online discussion of the Atkins Report mentioned earlier. The CART CMS discussion forum posting that follows was authored by a BSC computer science major participating in the course where she describes in her own words and references an initial definition of cyberinfrastructure:

> ***Atkins Report Class Discussion, by Cynthia Barnes***
>
> “Cyberinfrastructure was a brand new term to me when this class started and though I could speculate as to its meaning I was looking for a good definition while reading the Atkins Report (1). I picked up the meaning but didn’t really find what I wanted. I did, however, find a definition in a report called "Our Cultural Commonwealth." They state that, "cyberinfrastructure is meant to denote the layer of information, expertise, standards, policies, tools, and services that are shared broadly across communities of inquiry but developed for specific scholarly purposes (5)". So it’s something more specific than a network but more general than a tool. The base technologies of cyberinfrastructure are computation, storage, and communication. Without these integrated electro-optical components, cyberinfrastructure wouldn’t be possible. The whole point, as I understand it and as is stated in the Atkins Report, is to revolutionize what people can do, how they do it, and who participates by enabling them to share and collaborate over time and over geographic, organizational, and disciplinary distance. Without the base technologies there wouldn’t be much progress, time and money would be wasted. I think achieving the vision of the Advanced Cyberinfrastructure Program will be so beneficial to all fields of study that’s its worth the extra $1 billion annual budget. Like we discussed in class, when it comes to bioinformatics, the collaboration of cyberinfrastructure is needed to find out what is really going on in a living organism. This fact was backed up in Science in an article called “Cyberinfrastructure: Empowering a “Third Way” in Biomedical Research”, where it said, "Biomedicine is at the precipice of unlocking the very essence of biologic life and enabling a new generation of medicine. Development and deployment of cyberinfrastructure may prove to be on the critical path to obtaining these goals (6)".

The full article is available at https://www.vbi.vt.edu/article/view/554/1/15

At the beginning of our work with the learning modules, we took the following approach. The outbreak scenario was first introduced to the students. In this phase, the researchers of the CI-TEAM provided background information necessary for the students to understand the scenario and to develop strategies for investigations including identification of the pathogen, accessing the national network of bioinformatics resources, and suggesting control or preventive measures.

Specifically, the project required students to use multiple computational and web-based tools to discover the identity of the pathogen, given some preliminary biological data under the supervision of the faculty and in consultation with the CI-TEAM. First, students needed to learn about the DNA sequencing that generates a code for a portion of a pathogenic genome derived from an environmental sample. Next, the project required learning about how gene sequences are stored and how to access these sequences to test a hypothesis regarding the origin of the sample sequence. Students practiced what they had learned by becoming familiar with large, comprehensive repositories such as the NCBI database Gen-Bank, and focused bioinformatics resources such as, Patho-Systems Resource Integration Center (PATRIC), Pathogen Information (PathInfo), and Molecular Interactions Network (MINet) documents that are maintained at VBI or elsewhere on the Internet (He et. al., 2004, He et. al., 2005). Techniques for comparing and manipulating sequences were addressed to insure students could use similarity to a known genetic sequence to predict possible function for a sequence of interest. In the process, putative genes needed to have their sequences translated to protein sequences for functional features to be identified. Predictions of structural features of the protein, such as secondary structure and trans-membrane segments, gave the first clues as to what types of proteins were present. It was then necessary to characterize the family and use the sequences to generate a multiple sequence alignment to profile the family for subsequent investigation. Once family connections were made, the possible cellular role would be researched to assess whether the protein would be a potential candidate as a therapeutic target. Some students used molecular modeling of their proteins, where appropriate, during the investigation. Target selection was then made after several proteins were investigated using multiple data sources, including expression data where available. Although it was not utilized this time, subsequent experimental laboratory analysis could provide students with new information about the target that may be analyzed to generate possible leads in the search for drug candidates.

The students then worked through the scenarios taking on different roles in the team. Each team utilized different tools and addressed the outbreak question from a different perspective. The members of each sub-team then generated reports for their analyses and presented them to the other team members. The reports needed to contain enough information and to be presented in a clear fashion for other members to understand the information and for the entire group to agree upon a conclusion. It was necessary for students to cooperate to achieve the best results. Finally, the students assembled the information, synthesized interpretations and conclusions, and communicated the results to the live and online community. As referenced earlier link, https://www.vbi.vt.edu/article/view/554/1/15 one student described her involvement to the local media, and the interview was posted for the government bio-defense agencies such as DOD, NIAID, and the scientific experts in this process.

Communication and Interaction using CART Course Management Service (CMS)

The CMS became the online framework for the course called COSC 490 – Cyberinfrastructure. The CMS software and systems developed by CART at BSC allowed for full online course administration and access to syllabus, outline, surveys, quizzes, testing, reading material, chat, instructor collaboration and general remote student communication as substantiated in the article referenced above.

Supplemental Materials Used and Online Course Development

The course modules were designed for flexibility for the BSC professors, students, and the VBI team. As a test deployment, we needed to provide multiple instructor and assessment officer privileges and the CMS accommodates this requirement seamlessly. To supplement the modules, BSC professors enhanced understanding through live and online walkthroughs and interpretations. The online course was developed in direct and ongoing consultation and monitoring with the CI team. It followed an intensive week-long summer session where we worked together to design the fall semester course.

Student Demographics Important Factor in Course Implementation

For the first offering of the course, seven students were initially enrolled. These students were primarily computer science and engineering technology majors. One biology student enrolled; we had originally envisioned a more even distribution. This tended to steer the course more towards a cyberinfrastructure emphasis than a bioinformatics inclination.

Teaching and Learning Partners

The project team at VBI consists of the following professionals: (1) Dr. Oswald Crasta, Co-Director of Cyberinfrastructure Group, Computational Biology at VBI; (2) Dr. Susan Faulkner, Project Director, Education and Outreach at VBI; (3) Dr. Stephen Cammer, Senior Bioinformatics Scientist at VBI, (4) Dr. Daphne Rainey, Bioinformatics Scientist at VBI and (5) Dr. Betsy Tretola, Associate Director, Research and Assessment, Institute for Distance and Distributed Learning at Virginia Tech. The faculty members from BSC include: (1) Mr. Lionel Craddock, Professor of Computer Science; (2) Dr. Martha Eborall, Professor of Biology; (3) Mr. Frank Hart, Dean of Engineering Technology and Computer Science, (4) Dr. Lewis Foster, Professor of Physics; and (5) Mr. Bruce Mutter, Director of the Center of Applied Research and Technology.

Working together, the faculty at BSC and the VBI CI team designed and implemented a project-centric course focusing on CI for Life Sciences. The CI-TEAM Demonstration Project adopted three main phases to include (1) design of the CI course, (2) implementation of the course by the BSC faculty with participation by the VBI CI team, and (3) evaluation of the CI-TEAM course by the faculty and students at BSC under the guidance of Dr. Betsy Tretola.

Assessment and Lessons Learned

During the course of 2006, interaction between faculty and students has deepened our awareness of several key factors that have impacted implementation of the CI course at BSC. A thorough summative assessment was document and observations from faculty and student were assembled in the form of lessons learned.

Pre- and post-course online assessment surveys designed to measure the effectiveness of the proposed activities in the NSF CI-TEAM Demonstration project were administered to the BSC students enrolled in the CI course during fall 2006. Seven students (three females and four males) enrolled in the CI course. As referenced above, the class was primarily comprised of computer science majors. Five students reported having completed more than 13 hours of computer science courses. Two students reported completion of six hours of engineering courses, three students reported that they had completed six hours of math and statistics, and three students indicated that they had completed 13 or more hours in math and statistics.

While seven students completed the pre-survey, only six students completed the post-survey. One student, a physician, did not complete the course. On the pre-course survey 29% or two students agreed that they knew CI and multidisciplinary scientific concepts that they can effectively discuss with scientists, peers, and professors. On the post-course survey 83% or five students agreed that they knew CI and multidisciplinary scientific concepts that they could effectively discuss with others. On the pre-course survey 72% of the students indicated that their comfort level in working on a multidisciplinary team was good or excellent. On the post-course survey 83% indicated that their comfort level was good or excellent in working on a multidisciplinary team. These findings support the effectiveness of the course in meeting learning objectives related to communications and working within transdisciplinary teams to solve complex scientific problems. On the pre-course survey 57% or four students agreed or strongly agreed that they planned to pursue an advanced degree in a bioinformatics field. On the post-course survey 50% or three students indicated that they planned to pursue an advanced degree in the bioinformatics field.

In summary, the course helped students clarify their goals relative to doing CI research and seeking advanced degrees in preparation for work in a bioinformatics field. Students also confirmed that the course enabled their knowledge of CI, as well as, their ability to effectively discuss CI information with scientists, peers, and professors. Further, students indicated that the course increased their comfort-level in working on multidisciplinary teams.

VBI delivered a pre-course training workshop to BSC faculty to enable the design of the course and also provided overall support during implementation of the course in the fall 2006 semester. Surveys of BSC faculty related to the effectiveness of the pre-course training workshop and the overall support also were favorable.

A collection of our observations and recommendations are as follows:

Overview of BSC observations

- The project-centric course design did sufficiently engage students and ensured that a new degree of knowledge was obtained from the different discipline areas outside their respective comfort zones.

- The use of CI platforms, particularly the powerful web-based tools mentioned earlier, did instill a desire within students to utilize similar systems in other cross-disciplinary domains.

- The small number of students in the class recognized a need to further pursue knowledge in the bioinformatics domain, particularly in biology, after having taken the course.

- The professors appeared to remain engaged and interested in pursuing further research in CI and other collaborations with bioinformatics researchers, as a result of having taught the demonstration course.

- Both the students and faculty subsequently expressed an interest in further contributing to the CI at VBI by participating in the internships, design and research in the development of middleware or application components, such as TB/PP visualization plug-ins, new data sources, and new analysis tools.

- Finally, a project-centric course can be designed to serve two or three different student populations and used to disseminate the concepts of CI in an online environment. There remains much work yet to be done in all parts of the course described above.

Multi-stage approach to Bioinformatics Education

During the development and delivery of the course, focus on the goal of achieving team problem-solving. A multi-stage approach to education would be to emphasize tool use, select a problem to solve, and to provide background as needed. The level of background knowledge given can be adjusted to fit student profiles. Focus first on the problem and slowly combine minimal background theory as needed so as not to overwhelm students, especially those coming from backgrounds such as mathematics, engineering, and computer science.

Student number and background

Arrange course offerings to attract a mix of students from different backgrounds to give the true 'flavor' of CI interaction. When the student disciplines are relatively homogenous, several rounds through the modules will allow students the opportunity to play all the different roles and will increase the understanding of team interaction from different backgrounds. Greater application of the CMS could improve outreach, increase enrollment, and the diversity of backgrounds necessary for project-centric learning models.

Summary

We have described here the development and implementation of a CI course that was based on flexible project-centric learning modules. Along with these modules, supplemental materials, and the use of the CART CMS made for the successful implementation of bioinformatics education. The modules were easily adapted to incorporate various areas of genomic research and can be adjusted to diverse student backgrounds. The course can be adapted to several teaching formats from classroom to sustainable online implementations. CART at BSC is currently working to implement the course as an online course in the near future.

References

1. Atkins, DE, Droegemeier, KK, Feldman, SI, Garcia-Molina, H, Klein, ML, Messerschmitt, DG, Messina, P, Ostriker, JP and Wright, MH. (2003) Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure. 3 Feb. 2003 (http://www.communitytechnology.org/nsf_ci_report/)

2. Greene, K. and S., Donovan. (2005) Ramping Up to the Biology Workbench: A Multi-Stage Approach to Bioinformatics Education. Bioscene 31(1): 3-11.

3. He, Y., R. R. Vines, A. R. Wattam, G. V. Abramochkin, A. W. Dickerman, J. D. Eckart, B. W. S. Sobral (2004) PIML:the Pathogen Information Markup Language. Bioinformatics 21:116- 121.

4. Rainey, D., Faulkner, S., Craddock, L., Cammer, S., Tretola, B., Sobral, B.W., and O., Crasta. 2007. A project-centric approach to cyberinfrastructure education. TeraGrid 2007.

5. Our cultural commonwealth: The Report of the ACLS Commission on Cyberinfrastructure for the Humanities and Social Sciences, July 18, 2006

6. Buetow, K (2005) Cyberinfrastructure: empowering a “third way” in biomedical research. Science 308(5723): 821-824.

Acknowledgements

This work was supported by funding from the National Science Foundation (OCI-0537461 to O. Crasta).

Bruce Mutter[1], Daphne Rainey[4], Susan Faulkner[4], Frank Hart[2], Lionel Craddock[2], Martha Eborall[3], Lewis Foster[3], Stephen Cammer[4], Betsy Tretola[5], Bruno Sobral[4], and Oswald Crasta[4]

[1] Center for Applied Research and Technology (CART, Inc.) and [2]School of Engineering Technology and Computer Science, [3]School of Arts and Sciences, Bluefield State College, Bluefield, West Virginia USA

[4]Virginia Bioinformatics Institute and [5]Institute of Distributive and Distance Learning
Virginia Tech University, Blacksburg, Virginia USA