Building a culture of computing in the sciences using images as data within a community of practice

Tessa Durham Brooks[1], Raychelle Burks[1], Mark Meysenburg[1], Erin Doyle[1], and Chris Huber[1]

[1]Affiliation not available

August 27, 2021

Abstract

The Digital Imaging and Vision Applications in Science (DIVAS) program was built to improve the computational self-efficacy and skill of first- and second-year college students majoring in biological and chemical sciences. Our three-year pilot study showed that the program could be successful in both fronts. The scholars, faculty, and staff who participated formed a community of practice that became the heart of the DIVAS program. Through this community, we expanded access to the image processing workshop in collaboration with The Carpentries, supported faculty and secondary educators in developing computing modules for their classrooms, and created and staffed a "writing center for computing" on the host campus. Overall, the DIVAS program has sparked a local computing culture. DIVAS interventions and resources are freely available for adoption by other institutions. We hope to grow the community in a way that builds student access and opportunities and supports educators in the process.

Introduction

Data sets in natural, physical, and computer sciences are now so massive that their analysis, interpretation, and visualization often cannot keep up with the rate at which they can be routinely produced (Labrinidis & Jagadish, 2012). This data explosion means scientists across disciplines must work together to build tools, models, and visualizations capable of exploring that massive, complex, multi-faceted landscape. This cooperation will lead to the development of new tools that will allow us to observe aspects of our universe that have previously been hidden or beyond our reach. These new tools have the potential to transform the foundations of how we understand our universe analogous to the development of the microscope centuries ago (*Let Us Now Praise the Invention of the Microscope*, n.d.) .

Education and training within the natural and physical sciences do not sufficiently equip students with the necessary experience and skills to meet emerging data processing and analysis needs and experts across scientific disciplines are raising this issue (Lau et al., 2012; Lu et al., 2021; Skuse, 2019; Pevzner & Shamir, 2009). To this end, we created the Digital Imaging and Vision Analysis in Science (DIVAS) project.

DIVAS secured funds in 2016 from the National Science Foundation's Improving Undergraduate Science Education division (*NSF Award Search: Award 1608754 - Doane DIVAS: Digital Imaging and Vision Applications in Science*, n.d.). The program aims to engage novice learners (mostly first- and second-year college students) from biological and chemical sciences in computational thinking and coding using image data as a 'hook'. Images are widely used for diagnostics, phenotyping, and analytical measurements. They are also easy to obtain and novice learners understand them. When students analyze image data they engage in all of the aspects of computational thinking including recognition of a problem, analysis of solutions, design of a solution, and implementation and testing.

Of note, we saw the computational confidence and skills improve in students not normally well represented in computer science courses. Our pilot study indicated that the scaffolding of interventions that make up the program and the context of a community of practice were important elements contributing to the gains we observed (Brooks et al., 2021). The pilot study motivated us to see how broadly these results can be replicated. In this report, we discuss the key elements of the DIVAS project. We will also discuss what we have learned about fostering self-efficacy in computing more broadly across the natural and physical sciences.

DIVAS Project Overview

The DIVAS program targets first-year students enrolled in introductory biology and chemistry courses. The program name, its leaders, and the recruiting strategies used resulted in a high percentage of individuals who identify as women joining the program (76% of all program scholars) [8]. The program begins in the spring with a one-credit seminar and continues five to eight weeks into the following summer. The program ends with a capstone seminar the following fall or spring. From the first DIVAS seminar, students join a community of practice that will support them throughout the rest of the program. In that seminar, scholars learn about images as a form of data; do basic coding; meet professionals who use coding in their work; and explore working environments where coding occurs.

The summer begins with an intensive, one-week coding workshop which is also open to community members. The workshop devotes two days on basic Python operations, code version control using git, and bash shell functions. The following three days focus on image processing using Scikit-image libraries. We seat participants in pairs to encourage them to cooperate in their work. To increase engagement, participants are presented with a problem at the beginning of each portion of the workshop. By the end of the workshop, everyone has written scripts to count the number of colonies on a plate and to track the progression of a titration over time using a handful of basic image processing functions (Fig. 1).

Figure 1: Erin Doyle presents the challenge of tracking the progress of a titration to coding workshop participants.

Following the coding workshop, there are two two-week sessions of pair programming. We give the group a common challenge at the beginning of each session. Scholars discuss the problem, then together decide on the group's goals for the session. We teach scholars how to do pair programming, then divide them up to devise and implement their own strategy for solving the problem. Each day, the group holds a stand-up meeting to report on progress and discuss daily goals. At the end of each week, scholars push their code to a common repository and the teams print out and annotate each other's code for review. In the first week, code reviews help teams set their goals for the following week. In the second week, reviews help them wrap up their code. In the second two-week session we reassign pairs and repeat the process.

The program ends with another one-credit seminar in which scholars revisit the DIVAS repository to clean up and annotate code, and make adjustments to the coding workshops as needed based on feedback. They also meet with incoming scholars to welcome them to the community and offer support. Scholars also learn about parallelization and gain experience with parallel programming. Parallel programming is relatively straightforward to apply with the image processing scripts they've already written.

Pilot study outcomes

The DIVAS pipeline was tested on three cohorts of up to six scholars over three years. Seventeen scholars participated, 14 of which attended Doane University, a private liberal arts college in Nebraska. The other three scholars came from our partner campus at St. Edward's University, a private liberal arts college as well as a minority-serving institution, in Austin, TX. Most scholars identified as women (76%) and were in their first year of college (82%) majoring in biological or chemical fields (82%). Overall, we saw self-efficacy in computing increase by 34% on average as well as statistically significant growth in all the computational thinking skills we measured (recognize the problem, analyze solutions, design a solution, implement a solution) (Brooks et al., 2021). Self-efficacy grew even while interest in pursuing careers using computational skills did not and coding tasks became more challenging. Our previous publication includes details of the pilot study, including resources and information about each of the DIVAS program interventions (Brooks et al., 2021).

Dissemination of the image processing workshop

Image processing has become routine in studies of atomic, molecular, and cellular dynamics, those that associate genomic elements with phenotypes of interest, in breeding programs, and in a variety of monitoring and modeling in fields such as agriculture, ecology, and drug development. This increased demand within the scientific community for image processing skills has led us to turn the image processing elements of our workshop into a Data Carpentry lesson (*Data Carpentry*, n.d.; *Image Processing with Python*, n.d.). Data Carpentry supports community-driven development of domain-specific lessons to meet research training needs.

The lesson is still in the early adoption process. The workshop originally used OpenCV libraries. Community members converted it to using Scikit image libraries, which are much easier to implement across a range of platforms and environments. The lesson has been tested at three research institutions in the United States and Germany. The lesson assumes basic knowledge of Python, git, and bash and covers the basics of image processing, including image representation; creating histograms; blurring and thresholding; drawing and masking; edge detection; and object segmentation using connected components analysis. The two challenges that DIVAS scholars work on in this portion of the workshop are also available in this lesson. DIVAS project investigator Mark Meysenburg currently maintains the Data Carpentry lesson. As others in the Carpentries community use it and additional needs are identified we hope to see this lesson adapt to meet those needs.

Peer teaching

The Computing Center for the Liberal Arts. An important consequence of the pilot study was the creation of a broader community of students with computing skills on the Doane University campus. No longer siloed into specific departments and programs, students who once may not have interacted with each other academically were now connected through common interests and skills.

The DIVAS team recognized that scholars were broadening their community of practice to include peers who needed to build their own computational skill, as well as peers with more expert knowledge that could provide support. To help build this community further, the DIVAS team created a "writing center for computing" at Doane called the Computing Center for the Liberal Arts (CCLA). The CCLA is a place for anyone within the Doane community to get feedback and help with any computing project, from setting up an Excel spreadsheet to research using Doane's supercomputer, Onyx.

Using funding from the National Science Foundation, the team hired a center director whose duties included packaging CCLA program elements for broad implementation in a "computing center in a box" (*NSF Award Search: Award 1924094 - CyberTraining: Pilot: Institutional Cyberinfrastructure Training Center in a Box*, n.d.). Like writing centers across the nation, the CCLA is fueled by peer-to-peer support. Peer instructors develop short tutorials and guides on common computing needs, provide support via a Slack workspace and email address, and provide face-to-face support in the learning commons. The CCLA is now in its third year. A quarter of DIVAS scholars have served as peer consultants and one former scholar is the current training manager for 11 new consultants. Individuals majoring in six different disciplines have utilized its services. The center is growing steadily as the campus community becomes more familiar with its goal. The CCLA team has reached out to the humanities and social sciences departments to support their computing needs as well.

Introductory biology and chemistry courses. The DIVAS team has brought computational thinking through image analysis into introductory biology and chemistry courses in two ways. First, we designed an image processing module for general chemistry students to investigate the hydrophobicity of materials by measuring the contact angle of a drop of water upon them. Students take pictures of the water droplets using their smartphones and then use ImageJ, an open-source and widely used image analysis and processing tool (*ImageJ*, n.d.), to manually measure the angle of incidence from each image. This module has now been used for over five years.

Second, in an inquiry-based introductory biology course, mostly first-year students used a Google Colab notebook written by former DIVAS scholars to analyze images from a system to measure bacterial movement toward molecules (chemotaxis). A drop of agarose containing a test molecule or saline (control) is added to the center of each well in a six-well plate. The solid agarose droplet is then surrounded by stained *E. coli* in saline. Images are taken over time using flatbed scanners. Students first use ImageJ to develop their own strategies for measuring the change in cloudiness around the agarose plug over time (an indication that chemotaxis has occurred). Completely novice coders then review the code in the Colab notebook by drawing out a flowchart of what they see the code doing. When students were unsure of how the code was functioning, they tried changing part of it to see what effect it had. As a group, the class added clarifying annotations and create a full code map. From there, students used this code and modified parameters as needed to measure the change in cloudiness around the agarose plugs in their own well plates. This intervention is now in its second year.

Teacher training workshop. Finally, Nebraska high school teachers used another Google Colab notebook written by former DIVAS scholars last summer. Scholars wrote this code to measure the height and density of invasive grass growing in pots. The grass system provides a variety of angles to engage students in inquiry-based learning and to learn about asking questions, designing experiments, and analyzing and representing data. Teachers explored the pros and cons of manual measurements. They then studied the image-based approach to taking similar measurements by creating their own code maps as they worked through the Colab notebook. As teachers worked through the code, they identified other ways to take the same measurements and how to test the code to confirm it was functioning. Most of these teachers had never coded. Nevertheless, within a couple of hours, they learned to use the notebook. Further, they could evaluate it for use in the classroom both as a measurement tool and as a source of inquiry-based curriculum.

Broadening the alliance

Over the three years of the pilot study, we have learned a lot about how to support self-efficacy in computing early in a student's college career. The pilot data and anecdotal experiences alone might inspire other schools to try similar approaches. However, we still need to test the DIVAS program interventions across institutional types and different student populations. Broader implementation of all or part of DIVAS program elements will also help us identify its most critical elements. Teams wanting to implement a more streamlined program could implement just those elements with success. A broader DIVAS alliance will provide additional opportunities for students to collaborate, build their skills, and strengthen the community of practice. We also hope to expand the teacher training workshops, including both secondary and undergraduate educators. These workshops would use the Training of Trainers (ToT) model, empowering educators to integrate DIVAS interventions into their classrooms and research labs.

Discussion

We have long needed to find ways to infuse computational thinking, coding, and the use of scientific software into natural and physical science undergraduate education. Our experience with the DIVAS project, our pilot study, and the additional opportunities it fostered suggest that this can be done in many environments friendly to a community of practice approach. We have seen how a community can change the way students view computing from a specialized, esoteric skill to a set of tools anyone can learn to use. We have seen that novice learners can learn to use computational tools to solve problems relevant to their disciplines, gaining confidence in computational skills, and highly desirable workforce skills. As DIVAS program elements are adopted at other institutions, we will see this impact more clearly.

Tessa Durham Brooks is an associate professor of Biology at Doane University in Crete, NE. She completed her bachelor's in Biochemistry at the University of Nebraska and her Ph.D. in Cell and Molecular Biology at the University of Wisconsin. She teaches courses in introductory biology, physiology, and a third-year seminar on making meaning. Her teaching focus is to promote a sense of belonging in order to enhance learning in inquiry-based and flipped classroom environments. Durham Brooks and her undergraduate team explore phenotypic responses of plants to environmental stimuli and the effects of stress in early development on later growth phenotypes. She has an interest in developing infrastructure and academic experiences that promote computational and quantitative self-efficacy of undergraduate students in the natural sciences. Contact her at tessa.durhambrooks@doane.edu.

Raychelle Burks is an Associate Professor of Chemistry at American University in Washington, DC. Her lab research team is focused on the development of colorimetric and luminescent sensing systems with integrated image and chemometric analysis for forensic applications. She is on the leadership team of the Digital Imaging and Vision Applications in Science (DIVAS) project and DIVAS Scholars research advisor. Beyond the bench, Dr. Burks is a popular science communicator appearing regularly on TV, radio, podcasts, and print outlets. Central to Dr. Burks' research, teaching, and service is the central tenet that equitable, diverse, and inclusive practices both respect people and produce scientific outcomes of greater integrity. She is a member of several local, national, and international committees, task forces, and projects focused on social justice and STEM. Contact her at burks@american.edu.

Mark Meysenburg is a Professor of Computing at Doane University, a small liberal arts university in Crete, NE. He teaches the university's programming sequence, networking, and cybersecurity courses. Mark's research interests are varied and eclectic, including evolutionary computation, machine learning, robotics, and computer vision. Mark also teaches an annual first-year seminar course, using intense role-playing games to teach the history of science. Contact him at mark.meysenburg@doane.edu.

Erin Doyle is an associate professor of Biology at Doane University in Crete, NE. She completed her bachelor's degree in Applied Mathematics at the University of Tulsa and her Ph.D. in Bioinformatics and Computational Biology at Iowa State University. Her teaching focuses on helping students understand the importance of mathematics, computer science, and statistics to modern biology, and supporting positive student experiences in these areas. She teaches courses in introductory biology as well as upper-level electives in genetics and bioinformatics and computational biology. Students in her undergraduate research lab use computational approaches to generate experimentally testable hypotheses and use experimental results to develop and refine computational models of biological processes. Recent projects in her lab have focused on the identification of disease susceptibility genes in rice plants and functional characterization of bacteriophage genes. Contact her at erin.doyle@doane.edu.

Chris Huber is an assistant professor of Chemistry at Doane University in Crete, NE. He completed his bachelor's in Chemistry at the University of Wisconsin-La Crosse and his M.S. and Ph.D. in Chemistry at the University of Minnesota. He teaches courses in general chemistry, analytical chemistry, and physical chemistry. Dr. Huber centers his teaching style on delivering course material using an array of pedagogical strategies to match students' different learning styles. Dr. Huber's scholarly work centers on developing sensitive spectroscopic sensors capable of detecting ultra-low (sub ppm) concentrations of blood toxins. Contact him at chris.huber@doane.edu.

References

Challenges and opportunities with big data. (2012). *Proceedings of the VLDB Endowment*, *5*(12), 2032–2033. https://doi.org/10.14778/2367502.2367572 https://www.smithsonianmag.com/science-nature/what-we-owe-to-the-invention-microscope-180962725/. https://www.smithsonianmag.com/science-nature/what-we-owe-to-the-invention-microscope-180962725/

Computer Science: The Third Pillar of Medical Education. (2012). *Creative Education*, *03*(06), 807–810. https://doi.org/10.4236/ce.2012.326120

Future directions of chemical theory and computation. (2021). *Pure and Applied Chemistry*, *0*(0). https://doi.org/10.1515/pac-2020-1006

The third pillar. (2019). *Physics World*, *32*(3), 40–43. https://doi.org/10.1088/2058-7058/32/3/33

Computing Has Changed Biology—Biology Education Must Catch Up. (2009). *Science*, *325*(5940), 541–542. https://doi.org/10.1126/science.1173876

https://www.nsf.gov/awardsearch/showAward?AWD_ID=1608754&HistoricalAwards=false. https://www.nsf.gov/awardsearch/showAward?AWD_ID=1608754&HistoricalAwards=false

Hacisalihoglu, G. (Ed.). (2021). Digital imaging and vision analysis in science project improves the self-efficacy and skill of undergraduate students in computational work. *PLOS ONE*, *16*(5), e0241946. https://doi.org/10.1371/journal.pone.0241946

https://datacarpentry.org/. https://datacarpentry.org/

https://datacarpentry.org/image-processing/. https://datacarpentry.org/image-processing/

https://nsf.gov/awardsearch/showAward?AWD_ID=1924094&HistoricalAwards=false. https://nsf.gov/awardsearch/showAward?AWD_ID=1924094&HistoricalAwards=false

https://imagej.nih.gov/ij/index.html. https://imagej.nih.gov/ij/index.html