

Chapter 29

Building a Data Science Program Through Hackathons and Informal Training in Puerto Rico



Patricia Ordóñez Franco, María Eglée Pérez Hernández, Humberto Ortiz-Zuazaga, and José García Arrarás

Abstract With the growth of data in a plethora of fields ranging from agriculture to medicine to finance, data science is quickly becoming one of the most in demand professional careers of the decade. However, only a handful of minority serving institutions in the US have a course much less a formal program or certification track in data science. This paper highlights a solution at a public minority serving institution, which is in a hiring freeze, to create an interdisciplinary data science program using local resources through both formal and informal training and hackathons in collaboration with top research institutions and industry leaders locally and abroad in data science.

Keywords Big data education · Data science education · Diversity · Interdisciplinary studies · Teaching analytics · Hackathon · Data science workshop · Collaboration

Learning Objectives

- Describe the importance of data science for every community,
- Give examples of effective methods for creating a data science program in your community,
- Define what is a hackathon,
- Extend and adapt the model of using hackathons and informal training to the needs of your community to build a data science program.

29.1 Introduction

The University of Puerto Rico Río Piedras (UPRRP) is a top biomedical research institution and one of the top producers of Hispanic Ph.Ds in Science and Engineering in the US (<https://www.nsf.gov/statistics/2018/nsf18304/data.cfm>, Table 9), yet it

P. Ordóñez Franco (✉) · M. E. Pérez Hernández · H. Ortiz-Zuazaga · J. García Arrarás
University of Puerto Rico Río Piedras, San Juan, PR, USA
e-mail: pattordonez@gmail.com

lags in computational science. While the UPRRP has an undergraduate computer science department, students who wish to study computational science in graduate school must do graduate studies in applied mathematics and take undergraduate programming courses or learn to code on their own. Yet, due to the increase in data in the natural sciences, there is a demand for scientists who can create hypothesis and then applying data analysis to extremely large data sets (often referred to as Big Data) to derive knowledge (South Big Data Innovation Hub 2018). For this task, faculty and students must learn how to manipulate and extract data from databases, as well as apply statistics and/or machine learning and infer knowledge from the results. They also need to work and communicate with people of different backgrounds because of the interdisciplinary nature of the tasks. These skills are attributed to a data scientist.

Unfortunately, the financial crisis in Puerto Rico has forced the university to be in a hiring freeze, making it impossible to hire a newly trained data scientist. The Increasing Diversity in Interdisciplinary Big Data to Knowledge (IDI-BD2K) is a program that takes an interdisciplinary approach to developing an undergraduate data science program at the UPRRP through the informal teaching of faculty and students. The use of innovative hackathons and workshops combined with faculty development through the IDI-BD2K has created community and helped to develop the field of biomedical data science on the island. The process is being facilitated through collaborations with a former alumnus of the university who is a leader in data analysis in the life sciences at one of leading high research institutions in data science today.

29.2 Relevant Background

29.2.1 *Bridging the Data Divide*

Many speak of the increasing digital divide in education; however, there also exists an increasing data divide among institutions of higher education in the United States of America and that divide is most evident in Data Science. To our knowledge, there is not one Historic Black University or College, or Tribal College in the United States that has a Data Science program or track and only a handful of Hispanic Serving Institutions that do. Few public, rural universities, 2-year colleges, and community colleges do either (South Big Data Innovation Hub 2018).

In 2013 the National Institutes of Health (NIH) created the training program Big Data to Knowledge (BD2K) to increase the workforce in Biomedical Data Science (Dunn and Bourne 2017) by creating national Big Data to Knowledge Centers (BD2K Centers). The IDI-BD2K is one of the diversity projects linked to the BD2K initiative, as described in Canner et al. (2017).

The IDI-BD2K program is an NIH funded program to increase diversity and it aims to train undergraduate students in data science to be able to participate in BD2K ongoing research at three BD2K Centers during the summer. This summer experience

would be key to their training and to their choice of future careers. However, the BD2K program can serve a dual purpose by also developing and/or strengthening data science initiatives at the home institution. In our case, we have used the UPRRP BD2K diversity initiative, named IDI-BD2K, to serve as the training initiative in BD2K.

29.2.2 Increasing Diversity in Interdisciplinary Big Data to Knowledge (IDI-BD2K)

The IDI-BD2K program is focused on recruiting undergraduate students early in their sophomore year. Students will pursue specific course sequences depending on their major built on existing courses from other disciplines, so that by their junior year they have attained complementary levels of knowledge in math, statistics and computing. Interdisciplinary cohorts will then converge in a course sequence on Biomedical Big Data. Biomedical Big Data I (BBD I) is based on the series of MOOCs developed by Rafael Irizarry of Harvard University (an alum of the UPRRP) named Data Analysis for Life Sciences in R (<https://www.edx.org/xseries/data-analysis-life-sciences>) and is focused more on the statistics aspect of data science. BBD II is a course we created to convert students from all disciplines into data scientists using DataCamp with Python (<https://www.datacamp.com/>). The course was developed after one of authors was invited to spend 6 weeks at Facebook in the data analytics group to better understand the requirements of data science for industry. She interviewed several minority employees to understand how they had arrived at Facebook and what they would include in a data science course to be successful at Facebook.

Selected students then attend summer internships at participating BD2K centers at Harvard University, the University of Pittsburgh and the University of California Santa Cruz. The following fall students would participate in interdisciplinary undergraduate research projects with local mentors in a capstone course. Throughout the program, students attend workshops, seminars, hackathons, and meals where they receive informal mentoring and training in biomedical data science from prominent biomedical data scientists, develop professional skills and are inspired to be successful data scientists (See Fig. 29.1).

Training opportunities for affiliated faculty members include workshops, seminars and hackathons as participants and/or mentors. The project also sponsors short summer research experiences, workshops and other activities at the collaborating BD2K centers, either for training or for establishing research collaborations.

29.2.3 Challenges

There have been many challenges from the beginning of the grant's implementation, including having to justify to administration the scientific merit of a hackathon for faculty, students and the community in the opening event. However, as mentioned

Year	Terms	Science majors	Comp. Sci. majors	Math majors
1	S1	QUIM3001	CCOM3030	CCOM3030
	S2	BIOL3101	CCOM3033	CCOM3033
2	S1	MATH3026	CCOM 3034	CCOM3034
	S2	CCOM3030	QUIM3001	(MATH4031)
Summer		Cohort activity- One week workshop		
3	S1		BBD 1	
			MATH5001	MATH5001 QUIM3001
3	Winter	Cohort activity		
	S2		BBD 2	
			(CCOM5050)	MATH5002 BIOL3101
Summer		10 week internship at BD2K Centers		
4	S1		Capstone I	
	Winter	Cohort activity		
	S2		Capstone II	

Fig. 29.1 Timeline of courses for Science, Computer Science and Mathematics majors. The red courses are the ones that exist in another discipline but are not required. The ones in parenthesis exist in the major as electives and the blue ones are to be created

before, different from other training programs that we have administered there are three major challenges that are derived from creating an interdisciplinary and intersectoral training program and they are as follows:

29.2.3.1 Creating a Cohesive Student Cohort

Our experience with other training programs has shown that it is essential to build a cohesive student cohort that provides support among themselves and that nurtures the learning environment to help students advance even with minimal input from mentors. Other federally-funded student training programs provide a stipend or fellowship that helps not only in attracting students, but also in coalescing students around a set of program required activities. Our original aim was to have students continue working with local mentors following their return from their summer at the BD2K institution. However, the NIH-BD2K program would not allow financial support of this inhouse activity. Initially we obtained support funding from the Faculty of Natural Sciences, but the University financial crisis together with the aftermath of Hurricane Maria dried up the funding. We have been able to slowly build a student group that includes present and past students of the BBD courses and students who have attended the BD2K summer programs. This group of students is the main cohort around which activities are planned. These include workshops, seminars, and hackathons. Some of these students are involved in research with local mentors for either credit or as volunteers.

29.2.3.2 Biomedical Big Data Courses

Our initial plan was for students in our Program to take two Biomedical Big Data (BBD) courses during their third year, prior to their participation in the BD2K summer experiences. To be prepared for the BBD courses, students need to have taken a course in Statistics and a course in Computer Science in their first two years. Though this sounded like a simple, straightforward plan, it has been difficult to establish. The main problem has been that students tend to take these courses later in their university years, and it has proven difficult to convince them and the faculty otherwise. Thus, by the time students become interested in developing their data science proficiency, they are usually well advanced in their academic years. This implies that many of the students that are selected to attend the BD2K programs are in their fourth (and sometimes fifth) year of studies. Some of them end up taking the BBD courses once they return to the University. To attack this problem, we have now developed a strategic plan of actively going after students in the Statistics and Computer Science courses to make them aware of the BD2K opportunities and the need to have taken the basic courses to be able to apply in their junior year.

29.2.3.3 Student Recruitment

The difficulty in moving students through our planned course sequence also decreased the number of students that had the required expertise to participate in our collaborator's BD2K summer program. Like in many other situations where a new program is being created, some flexibility is needed. In our case, we were able to identify populations of students that had the required expertise that our partners requested (i.e. computer programming, statistics...). Thus, advanced undergraduate students from the Mathematics Department and from the Computer Science Department were attracted to our Program and were offered a slot in our collaborative arrangements. This alternative source of capable students has provided us the time and space to sort out and fix our problem of recruiting a younger cohort to our BBD courses and eventually to the BD2K summer experiences. We also decided to use interdisciplinary hackathons as a recruitment tool.

29.2.4 *What Is a Hackathon?*

Hackathons have often been characterized by intense competitions where mostly males congregate to work tirelessly for 36–48 hrs while eating unhealthily to produce a product. Such environments have been informally demonstrated to be uninviting to women (Williams 2014). However, hackathons have been found to be incredible networking opportunities that can lead to the creation of companies and other opportunities which women and URM's miss out on. Thus, the hackathons we have created have been focused more on building collaborations, promoting inclusion

and developing and presenting of the process to a solution rather than a prototype to develop professional and technical skills in participants. Furthermore, while the hackathons typically last 36–48 hrs, we strongly encouraged sleeping the first night and eating healthy meals in community. To encourage faculty and experts to participate, hackathons were held in conjunction with professional development activities and lunch was considered a networking activity where expert mentors could meet with the hackathon teams and offer advice. Sitting with new people was strongly encouraged if you were not receiving mentoring.

29.2.5 Barriers to Increasing Diversity in Biomedical Data Science

As an island, Puerto Rico must import much of its resources from abroad. The strong talent that is produced on the island is often recruited to the mainland. For the purposes of the United States, any talent from Puerto Rico is increasing diversity in Biomedical Data Science. Here we list a few of the barriers that exist in increasing Biomedical Data Science on the island expanded from Canner et al. (2017) as they pertain to Puerto Rico.

1. A lack of preparation not only in the focus areas of informatics, statistics, and biology, but also in their breadth of understanding of how these disciplines can be integrated (Greene et al. 2016; <https://www.kaggle.com/surveys/2017>). The University of Puerto Rico Río Piedras is a very traditional university where there is more encouragement for transdisciplinary research than interdisciplinary research. Transdisciplinary research occurs when two disciplines transcend each other to discover unexpected knowledge or create new approaches to solving a problem. Interdisciplinary research requires an integration of the disciplines in the search of solutions to complex problems (<https://blogs.lt.vt.edu/grad5104/multiintertrans-disciplinary-whats-the-difference/>). We are fortunate to have an Interdisciplinary Program in the College of Natural Sciences. Students, however, build their own program and the program is therefore more multidisciplinary than interdisciplinary. Inadequate development of the professional and cognitive skills necessary for entrance to and success in graduate school. This is an especially significant hurdle, as many careers in biomedical big data require, at minimum, a Master's degree (Colbeck et al. 2001).
2. Limited opportunities for undergraduate research prior to graduation. This challenge is particularly acute at non-R1, four-, and two-year institutions where faculty-led opportunities to engage in research are limited (O'Donnell et al. 2015). In contrast, the UPRRP offers many research opportunities for undergraduate research in biomedical research. However, few of the opportunities require computational rigor.
3. While there is not a lack of understanding of the rigors and research culture of the biomedical field at UPRRP, and it does not conflict with personal cultural identity

of the university, there is evidence to indicate that there is a lack of understanding may lead to a lack of diversity in biomedical big data and discouragement for underrepresented groups to pursue biomedical research (Malcom et al. 2010). In Puerto Rico, the lack of understanding of biomedical data science and big data, and of the difference between biomedical informatics and bioinformatics has been a limitation.

4. An absence of exposure to innovative undergraduate level curricula that develop the skills and concepts relevant to the world of big data, while also allowing students to focus on specific sub-disciplines of this broad field (Greene et al. 2016; O'Donnell et al. 2015).

29.3 Methods: Developmental Activities Through Informal Training and Hackathons

Each of the activities described below are either a specialized hackathon or a workshop. In all cases, both faculty and students received informal training and were exposed to research or professional development. The purpose of these activities was to bring faculty and students together with industry to further knowledge of biomedical data science, big data, or health informatics and to create interdisciplinary and intersectoral collaborative teams to solve transdisciplinary problems. To our knowledge, this is a novel approach to spur innovation in and disseminate knowledge about interdisciplinary data science.

The event to kick off the IDI-BD2K program was a hackathon on health informatics. It ran parallel to a health informatics symposium to motivate the hackathon and thus the first day was in parallel. The second day the participants separated into three sessions: (1) the symposium in health informatics, (2) the workshop in biomedical data science, and (3) the hackathon.

29.3.1 Symposium of Health Informatics in Latin America and the Caribbean

The Symposium of Health Informatics in Latin America and the Caribbean (SHILAC) unites two main areas to facilitate the creation of technological tools in improving the quality of health in Latin America and the Caribbean: health and information technology.

For the first time, San Juan, Puerto Rico hosted this important event in which health professionals, hospital administrators, health service providers, scientific researchers, public health specialists, physicians, and technology developers participated to discuss and plan the creation of technological products to face the pressing needs of health in Latin America and the Caribbean. SHILAC 2015 was held on November 20–22, 2015, at the San Juan Marriott Resort in Condado, San Juan, Puerto

Rico. The hackathon was held in conjunction with SHILAC to attract mentors from industry, government and academia to one location.

The conference featured speakers of recognized prestige from Latin and North America, and included Leo Celi, M.D, of MIT, Carol Hullin, Ph.D from World Bank, Lucila Ohno-Machado, Ph.D from the University of California San Diego, and Juan Carlos Puyana from the University of Pittsburgh Medical Center. It also included panels from the Hospital Association of Puerto Rico, the Industrial Association of Puerto Rico, Ponce Health Sciences University, among others to bring together academia and industry from all parts of the island.

29.3.2 Biomedical Data Science Workshop

Concurrently with the first day of SHILAC, a Biomedical Data Science workshop was offered for underrepresented students from the US and Puerto Rico, sponsored by the Computing Research Association, special interest group in women (CRA-W). The workshop was also open for interested faculty. The workshop was targeted to novices on Biomedical Data Science. Speakers included Hector Corrada, Ph.D from the Center for Bioinformatics and Computational Biology (CBCB) at University of Maryland—College Park, Roger Mark, M.D, Ph.D from MIT Laboratory of Computational Physiology, Gabriel Kreiman, Ph.D from The Center of Brains, Minds and Machinery, Tyrone Grandison, Ph.D, Deputy Chief Data Officer (dCDO) at the US Department of Commerce. Some of the hackathon teams originated from this activity.

29.3.3 Hacking Health in the Caribbean

The hackathon named Hacking Health in the Caribbean was directed by the well-known group, MIT (Massachusetts Institute of Technology) Hacking Medicine (<http://hackingmedicine.mit.edu/healthcare-hackathon-handbook/>) and had mentors from the SANA group from the MIT Laboratory of Computation Physiology, known for its hackathons in global health (<https://www.tandfonline.com/doi/full/10.1080/03091902.2016.1213903?scroll=top&needAccess=true>). It was the first hackathon related to health in Puerto Rico as well as the first time that MIT Hacking Medicine performed a hackathon in the Caribbean and Latin America.

This event attracted teams of faculty and students as well as a few persons from industry and government that served as mentors. By the end of the first day, over 30 problems had been pitched and more than 10 teams were formed to develop projects to solve these problems using health informatics in a period of 3 hrs. Different from most hackathons, participants were required to get a good night's rest the first night and begin at 8 am the next day to do a 24-hr hackathon. On the second day, they ate meals and mixed with mentors at lunch and dinner and they received a healthy snack at midnight to re-energize. They had a room to themselves where the hacking

and mentoring occurred and where they were encouraged to do a pre-presentation after 5 pm for feedback. The final presentations occurred from 8 am to noon the third day and judges from industry, government and academia deliberated during lunch as a panel including a member from every team was interviewed by the Organizing Committee Chair to reflect on the event. Prizes for best papers and posters of SHILAC and prizes for the hackathon were given during lunch.

The winning teams from the hackathon included the projects:

1. First place: Digital platform to determine the level of risk in air quality for asthmatics and people suffering from allergies.
2. Second place: Tracking bracelet for patients with dementia.
3. Third place: Online resources for parents of children with autism.
4. Best Hack sponsored by AARP: A digital platform to connect caregivers of the elderly abroad with the aging community on the island.
5. Best Hack sponsored by Varmed Management Corp (tie):
 - a. Data Visualization of Super Utilizer data, a Puerto Rican group of undergraduates.
 - b. Visual Analytics of Super Utilizer data, a Colombian group of graduate students.

Faculty and students from the University of Puerto Rico (Humacao, Río Piedras and Mayagüez), Metropolitan University and the Inter-American University (Metro and Arecibo) as well as students from universities across the United States participated.

29.3.4 Healthcare Innovation Replicathon

The Replicathon took place on March 24–25, 2017 at Engine-4 in Bayamón to allow undergraduate and graduate students to experience a mentored opportunity to work on a collaborative project in Biomedical Data Science. A Replicathon is considered as a form of hackathon, but in this case, all the participants are working on the same problem, trying to reproduce the results of a biomedical data science research publication. Puerto Rico was the birthplace of this innovative event designed to train students to work in interdisciplinary teams consisting of students of Biology, Mathematics, Information Technology, Medicine, Public Health, Computer Science and Statistics among others. The objective of the event is to attract students to the computational and quantitative sciences and develop in them the skills of collaboration and critical analysis necessary to solve real problems in science.

Like a hackathon, a Replicathon requires students with programming skills to create real solutions using technology. Unlike a hackathon, all teams analyze two scientific manuscripts that arrived at two different conclusions about the same data and present their interpretations of the results. In a hackathon, the solution is usually done in the form of an App (a mobile or web application). Replicathon requires

interdisciplinary collaboration between experts in programming, data analysis, and content (genomics in this case) and the solution is presented as a Jupyter notebook for context experts and as a presentation for a panel of scientists and industry leaders.

After the welcome, the event began with a plenary talk by Dr. Tracy Teal about her company, Data Carpentry, which focuses on teaching introductory computational skills for the management and analysis of data for developing “efficient, shareable, and reproducible research practices” (Data Carpentry Mission 2018). Then the organizers explained what a Replicathon is and the goals and rules of the event for the participating students. Then Keegan Korthauer and Alejandro Reyes, doctoral students of the Laboratory of Biostatistics of Rafael Irizarry at Harvard University presented the problem. Interdisciplinary teams met during lunch and analyzed the data all afternoon and night on Friday until the afternoon of the next day.

29.3.4.1 Collaboration and Mentors

Mentors stayed with students during this time and the teams presented mentors their conclusions in the early evening. After incorporating suggestions from the mentors, the teams presented their final results to a panel of scientists and industry leaders not involved in the mentoring on Saturday morning. Meanwhile data scientists who were participating in a concurrent Data Carpentry training judged the Rmarkdown deliverable supporting the team’s stance. After the presentations, winners were decided by the Rmarkdown and presentation judges and the awards ceremony was held during lunch.

Mentors for the event came from the University of California Davis, Harvard University and #include <girls> , the largest student women’s organization in the computer field on the island. This event was the result of a collaboration established with the UPRRP researchers of the IDI-BD2K Project with Rafael Irizarry of the Rafalab of Harvard University (Puerto Rican biostatistician and former student of the UPRRP and considered one of the most influential biostatisticians in the United States; <http://www.elnuevodia.com/ciencia/ciencia/nota/cientificoboricuaentreloslo sfundeventsunited2012-2012296>) and Titus Brown of the Laboratory of Intensive Data Biology at the University of California Davis.

Meanwhile, faculty from Interamerican University Bayamon Campus, UPR Humacao, Mayaguez, Rio Piedras and private industry went through Data Carpentry Instructor Training led by Rayna Harris (UT Austin), Sue McClatchy (The Jackson Laboratory), and Tracy Teal (Data Carpentry). Data Carpentry Instructor Training presents instructors with research-based best practices for teaching data science to novices. Fifteen faculty and graduate students participated in this training workshop. Two of the participants subsequently completed the Instructor checkout and are qualified to teach Carpentry workshops. Here we combined informal technical training for faculty with a hackathon for students.

29.3.5 Additional Workshops

The UPR organized two Data Carpentry Workshops, one on Genomics and one on Ecology, from the 15 to the 18 of August, 2018. The Genomics workshop was sponsored by IDI-BD2K. Humberto Ortiz-Zuazaga from the IDI-BD2K, was one of the instructors for the Genomics workshop with Nelly Selem, a Ph.D student from the Universidad Nacional de México (UNAM). Four undergraduate students from UPR Río Piedras and a second Ph.D student from the UNAM were helpers: Eveliz Peguero, Sebastian Cruz, Israel Dilán, Kevin Legarreta González, and Abraham Avelar. Interest in the Genomics workshop was very strong, with 48 registrants. Space limitations required us to cap attendance at 35. Participants learned how to manipulate next generation sequencing data to see variants in a population of *E. coli*. To do this, they used cloud computing resources, logged in remotely, processed files on the command line, and wrote scripts to automate parts of the analysis.

29.4 Results

Attendance in events by the IDI-BD2K has been consistent and steady throughout the last three years (See Fig. 29.2). We estimate that we have reached approximately 900 people through our workshops. The mailing list currently has 128 people. We



Fig. 29.2 Students from private and public universities mentored by leaders in industry, health, academia, and government at our first hackathon as they innovate to solve health issues facing their communities

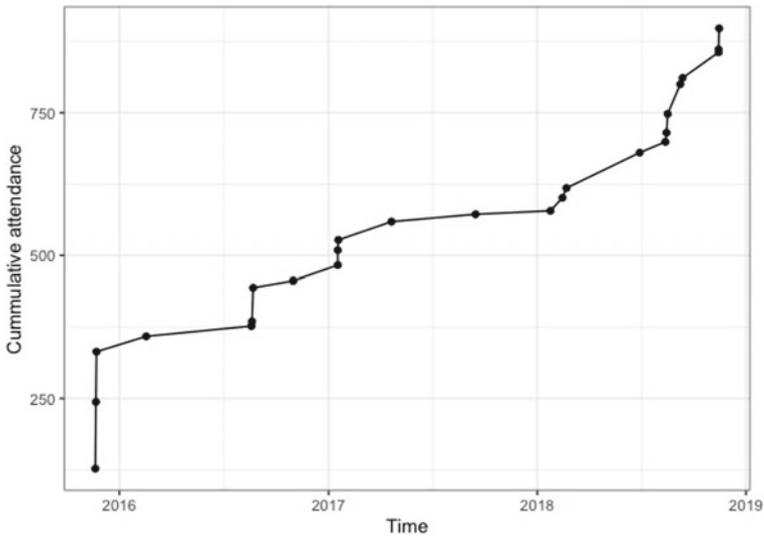


Fig. 29.3 Cumulative sum of attendance at IDI-BD2K activities

are reporting our results in terms of the lessons learned in facing our challenges and unexpected consequences. From the graph, it is easy to see the effects in 2017 of the student protests which closed the university for 70 days and of Hurricane María which closed the university for one month and caused a drastic decrease in attendance as a result of the massive power loss across the island, the longest power loss ever recorded in the United States of America and its territories (Fig. 29.3).

29.4.1 Establishing a Cohort of Students Interested in Big Data

One of our biggest challenges was to establish a cohort of students interested in Big Data that could benefit from the various activities offered by the grant and that at the same time could serve as the pool of students that could be developed and recruited to participate in the various summer programs. Our initial efforts focused on the students that returned from the summer programs, however, this group was too small, and their time left in the University (one year) too short to be able to form a stable cohort. Slowly, we extended our efforts toward “younger” students that were taking or had recently taken the basic Statistics and/or Computer Sciences courses. This combination of the younger students together with advanced students returning from their summer experiences has helped form a more stable cohort of students that participate in our BD2K activities and at the same time serve to recruit other students to the Program.

One of the major drawbacks of the BD2K program, when compared to other NIH-training programs, is the lack of financial support given to the students. Programs such as NIHMBS or NIH-ENDURE provide a stipend or fellowship to participating students. These students participate in research in their local universities and program activities during the academic year. This activity in itself strongly promotes the formation of a student cohort that shares experiences and academic goals. In addition, by providing this stipend, students are kept focused on big data research, continue their training/mentoring so that they advance toward graduate school, and serve as “unofficial senior mentors” to those students that are beginning in the program. This arrangement also keeps students from getting “computer-related” jobs outside the University (which often lead to students leaving Academia and entering the job market).

29.4.2 Establishing and Promoting Courses in Big Data

Our program established two courses in Big Data. These courses were directed at students in their junior year who were interested in Big Data analyses and were part of the training plan for students that would eventually go for summer experiences at the BD2K Centers. Initially, the courses attracted a very limited number of students, which required the support of the Math and Computer Sciences departments to keep the courses running with less than the minimum number of students required by the university administration. In trying to improve the situation we realized that the main problem was that our Big Data courses, as described, were labeled as “electives” for Math and Computer Sciences students, where the number of elective students can take in their programs are very limited. The second problem was a marketing problem. Simply by renaming the course to “Data Science” has resulted in a huge increase in student interest. Registration for the next semester course is five times larger than last year’s and the Department has had to put a limit on student registration.

29.4.3 Building Interest in Big Data Among Colleagues

The challenge of attracting colleagues to Big Data can be even more difficult than attracting students, particularly at institutions not readily involved in interdisciplinary work. Our strategy to attract faculty was to use the Program’s seminars and workshops. We tried to match the seminar’s topics to the faculty interests, where faculty could relate the seminar studies to their own work. Similarly, workshops were aimed at beginner levels where participants could be introduced to Big Data topics without feeling overwhelmed. Program participants come from three different departments (Biology, Mathematics and Computer Sciences) has also helped generate a level of enthusiasm that has served as an impetus to interdisciplinary activities and collaborations. A measure of our success in attracting colleagues into Big Data Science

and expanding the impact of our Program can be seen in next semester's *Topics in Modern Biology* course. This course is a required for graduate students in the Biology Program. The course topic changes every year, and the course is offered by visiting faculty that spend ~4 days at the University of Puerto Rico during which they provide a series of lectures, a research seminar and a workshop. Next semester's course topic is *Big Data in Biology: from genes to the biosphere*. The course has been organized by two professors from the Biology Department, and has the largest number of students registered ever. In addition, a section is being opened for interested undergraduate students to be able to take the theoretical aspects of the course.

29.5 Conclusions

There is little interdisciplinary and intersectoral culture in the College of Natural Sciences at the University of Puerto Rico Río Piedras. However, through the processes of these innovative, non-traditional, inclusive, interdisciplinary and intersectoral hackathons and activities, we have witnessed the growth of a biomedical data science community not just at the University of Puerto Rico Río Piedras, but throughout the island such as at the University of Puerto Rico Medical Sciences Campus, which started an online Data Science course last year. Given the current fiscal and hiring constraints, we aim to build a multidisciplinary Data Science program in the near future where by each discipline can create its own data science program using interdisciplinary courses in mathematics and computer science and culminate in an interdisciplinary capstone courses that will propel our campus into interdisciplinary data science research. In the future, we would like to expand these models to other countries in Latin American and the Caribbean which have similar constraints by leading similar events and conferences in these regions.

References

- Canner, J. E., McEligot, A. J., Pérez, M.-E., Qian, L., & Zhang, X. (2017) Enhancing diversity in biomedical data science. *Ethnicity & Disease*, 27(2), 107–116. <http://doi.org/10.18865/ed.27.2.107>.
- Colbeck, C. L., Cabrera, A. F., & Terenzini, P. T. (2001). Learning professional confidence: linking teaching practices, students' self-perceptions, and gender. *The Review of Higher Education*, 24(2), 173–191. Project MUSE, <https://doi.org/10.1353/rhe.2000.0028>.
- Data Carpentry Mission as stated in (2018). <http://www.datacarpentry.org>.
- Dunn, M. C., & Bourne, P. E. (2017) Building the biomedical data science workforce. *PLoS Biology*, 15(7), e2003082. <https://doi.org/10.1371/journal.pbio.2003082>, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5517135/>.
- Greene, A. C., Giffin, K. A., Greene, C. S., & Moore, J. H. (2016). Adapting bioinformatics curricula for big data. *Briefings in Bioinformatics*, 17(1), 43–50. <https://doi.org/10.1093/bib/bbv018>.
- Malcom, L. E., Dowd, A. C., & Yu, T. (2010). *Tapping HSI-STEM funds to improve Latina and Latino access to the STEM Professions*. Los Angeles, CA: University of Southern California.

- O'Donnell, K., Botelho, J., Brown, J., González, G. M., & Head, W. (2015). Undergraduate research and its impact on student success for underrepresented students. *New Directions for Higher Education*, 169(169), 27–38.
- South Big Data Innovation Hub. (2018). Keeping Data Science Broad: Negotiating the Digital and Data Divide Among Higher Education Institutions. https://drive.google.com/file/d/14l_PGq4AxOP9fhJbKqA2necsJZ-gdiKV/view.
- Williams, C. (2014). Why don't more women go to hackathons? (21 Jan 2014). Retrieved March 16, 2018, from <https://www.forbes.com/sites/quora/2014/01/21/why-dont-more-women-go-to-hackathons/#739a33716091>.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

