2023 RECOMMENDATIONS REPORT

Identifying Assets and Collaborative Activities to Support Student Success in Environmental Data Science at Minority Serving Institutions

Cite as: Alycia Crall, Sara Bolduc, Kendi Ho, Micaela Parker, Krystal Tsosie, Talitha Washington, Carrie Diaz Eaton, Claire Lunch, Kari Jordan, Camille Santistevan, Lou Woodley. (2023). Identifying Assets and Collaborative Activities to Support Student Success in Environmental Data Science at Minority Serving Institutions. Zenodo. https://doi.org/10.5281/zenodo.8231167

Note: This report is based upon work supported by the National Science Foundation under Grant No. 2135830. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Introduction and Goals

This report synthesizes the findings and recommendations from a project funded by the National Science Foundation (Award #2135830): *Identifying Assets and Collaborative Activities to Support Student Success in Environmental Data Science at Minority Serving Institutions*. The project brought together current and former faculty from Tribal Colleges and Universities (TCUs) and Historically Black Colleges and Universities (HBCUs) to discuss opportunities to empower and support equitable data science education across these institutions and co-create solutions that ameliorate the digital divide. Our collective vision is to ensure all students have the resources and skills required to be effective in applying data science throughout their careers, ensuring their competitiveness in the workplace where they will be active participants in using their knowledge and skills to influence policies and decisions that may directly impact their communities.

We focused specifically on the application of these skills to the environmental sciences, as there has been significant growth in large ecological datasets and environmental synthesis projects over the last decade. We aimed to co-create actionable recommendations for undergraduate environmental data science education at TCUs and HBCUs. This report represents the culmination of our project work, summarizing our activities and synthesizing the discussions we had among ourselves, our working group members, and the broader community. Our project goals were:

1. to identify the unique assets that these institutions bring to environmental data science education and the barriers to adoption of data science in teaching relevant courses;
2. to identify and raise awareness of the resources and other assets available through existing organizations that support equitable data science education;

3. to promote relationship-building among faculty and EDSIN partner organizations, forming the basis of a network for future community-informed resource sharing and co-creation of culturally-specific curricula with peer support; and
4. to co-develop this recommendations report for the enhancement of student success in environmental data science at these institutions.

For additional background, see our Project Justification in Appendix A.

Our Partners

To achieve these goals, the Academic Data Science Alliance partnered with The Carpentries, the Atlanta University Center Data Science Initiative, the Native BioData Consortium, NEON, and the RIOS Institute to build relationships with a network of colleagues at HBCUs and TCUs through monthly meetings and a series of mini-workshops. The convenings were facilitated and supported by the Center for Scientific Collaboration and Community Engagement (CSCCE). Analysis and synthesis of these many conversations, following the Strengths, Opportunities, Aspirations, Results (SOAR) model (Stavros and Cole, 2013), was led by Sara Bolduc Planning and Evaluation LLC (SBPE). SBPE also created and distributed a form to collect additional feedback from the larger community, allowing us to expand and enrich the project's insights. For full descriptions of the project partners, reference Appendix B.

The Process

The project was divided into three phases (Table 1). Phase 1 included members of the Project Leadership Committee (chaired by PI Parker), who began organizing and planning to support the working group (WG) meetings and workshops. The committee also recruited members to sit on each of the two WGs (TCU WG chaired by co-PI Tsosie; HBCU WG chaired by co-PI Washington). Each WG included faculty and leadership who brought diverse perspectives on the unique cultural landscapes at their institutions and the needs of their students (Appendix B).

Phase 2 engaged TCU and HBCU WG members in conversations surrounding the project goals through a series of meetings. During these meetings, each WG was charged with identifying topics of importance at their institutions with an emphasis on Goal 1. Phase 2 also included the distribution of a form to gather broader community input on project goals beyond the WGs with an emphasis on engaging students at these institutions. The team utilized the SOAR (strengths, opportunities, aspirations, and results) model as a framework for designing the feedback form and synthesizing the community feedback to align with WG activities (Stavros and Cole, 2013). It should be noted that the responses from the community to the feedback form were limited (n=21 responses), but taken together with the working group discussions, the insights described here can be thought of as a starting point for the larger community to build on for future projects, and while this document remains open for feedback.

In Phase 3, members of the Leadership Committee brought together both TCU and HBCU WG members for a final, combined half-day mini-workshop. The goals of this final workshop were to: share the synthesized outcomes identified in the SOAR analysis that emerged with contributions from the facilitated WG meetings and the community-wide survey.

Throughout the project duration, the leadership committee met approximately weekly to scope and plan (and debrief after) each working group meeting and the three workshops. Following the final workshop, the team and working group members took a break for the holidays. A small core writing team (Parker, Crall, and Bolduc) then resumed synthesis of the many discussions and SOAR feedback in the beginning of 2023 to draft this recommendations report. An initial version was circulated to the working group members in April 2023 for comment and feedback.

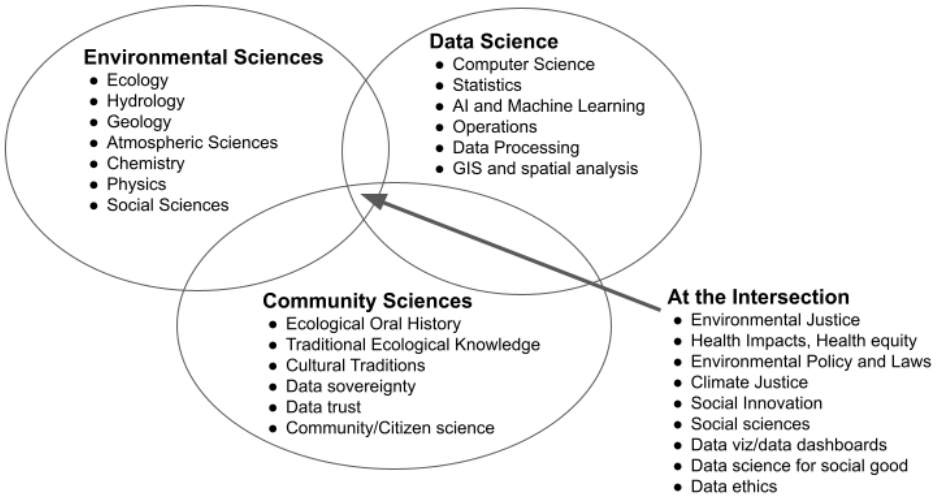
Table 1. Project phases, activities, and timeline.

Phase 1: Relationship Building & Planning	Leadership Committee Chair: Micaela Parker Outreach and creation of Working Groups Weekly leadership meetings (for the duration of the project)	September-October 2021
Phase 2: Assets and Barriers Identification	HBCU Working Group Chair: Talitha Washington Meeting 1 (11/21; 14 attendees) Meeting 2 (12/21; 10 attendees) Meeting 3 (02/22; 12 attendees) Meeting 4 (03/22; 14 attendees) TCU Working Group Chair: Krystal Tsosie Meeting 1 (11/21; 20 attendees) Meeting 2 (12/21; 14 attendees) Meeting 3 (01/22; 15 attendees) Meeting 4 (02/22; 15 attendees) SOAR feedback form creation and distribution Leads: Sara Bolduc, Kendi Ho, Alycia Crall	November 2021- April 2022
Phase 3: Co-Learning & Co-Creation	Joint Workshops (for community building, partner support & SOAR feedback) Lead: CSCCE Mini-Workshop 1 (04/22; 24 attendees) Mini-Workshop 2 (06/22; 14 attendees) Final Workshop (3) Community Engagement (10/22; 27 attendees)	April-October 2022
Analysis and Dissemination	Leadership Committee Chair: Micaela Parker Analysis of Working Group discussions and community feedback Leads: Sara Bolduc and Kendi Ho Reporting and Dissemination Leads: Micaela Parker, Sara Bolduc, and Alycia Crall	February - May 2023

Defining Environmental Data Science

Recommended actions to support student success in environmental data science first require defining what environmental data science means for each community and how student success will be measured. Borrowing from Figure 1 in "Data Science of the Natural Environment: A Research Roadmap" (Blair et al. 2019), both working groups created new clouds for environmental data science for their specific communities, where environmental data science lives at the intersection of three clouds: environmental sciences, data science, and community sciences. The two working groups then came together in one of our mini-workshops to create a visualization of environmental data science that works for both groups.

Figure 1. Defining Environmental Data Science at TCUs and HBCUs



Defining Success in Environmental Data Science

Defining student success was the next challenge. In both working groups, rich discussion and sharing of stories led to important insights, including that students should be given opportunities to define success for themselves and that student success can and should include positions outside of academia. And, when student success is discussed, greater attention needs to be paid to the pathways often taken by Indigenous students. For example, positions with tribal governments and tribal agencies are typically not presented in such discussions. Some common themes that surfaced as we discussed student success included:

- Inclusive curriculum and access to education
- Degree and career opportunities
- Improved student knowledge and skills
- Improved student confidence
- Improved student wellbeing
- Community collaboration/understanding

Overall Results of the SOAR Analysis

The team utilized a SOAR (strengths, opportunities, aspirations, and results; Table 2) model as a framework for leading WG discussions, listening to WG member feedback, designing and administering a community-wide feedback form, and synthesizing this work. The SOAR model focuses on "identifying strengths, building creativity in the form of opportunities, encouraging individuals and teams to share aspirations, and determining measurable and meaningful results" (Stavros and Cole, 2013, p.10). SOAR is rooted in appreciative inquiry and moves away from the traditional Strengths, Weaknesses, Opportunities, and Threats (SWOT). While SWOT seeks a balanced analysis of what *is* and *isn't* desired, SOAR has been reported to 1) shift the focus to the positive; 2) involve those who will impact and who will be impacted; 3) be more empowering; and 4) help sustain the momentum and nurture a "living" strategy (Silbert and Silbert, 2007).

Table 2. The SOAR framework

Strengths	Opportunities	Aspirations	Results
What are we already doing well to support students in environmental data science?	What could we be doing more of to support students in environmental data science?	What do we aspire to do (or become) to support students in environmental data science in the future?	How will we know we got there?

The summary of the rich discussions and feedback responding to these questions are presented below as they relate to TCUs and HBCUs.

STRENGTHS: What are we already doing well to support students in environmental data science?

Minority Serving Institutions (MSIs) have proven to be effective at recruiting, retaining, and preparing underrepresented students for the STEM workforce when compared to predominantly white institutions (NASEM 2019). Each MSI type (e.g., TCU, HBCU) brings unique strengths to support the success of their students, and those strengths are further facilitated by the individual institutions within each type. Tables 3 and 4 summarize the strengths in environmental data science education listed by working group members and form respondents by MSI type.

Table 3. Summary of responses listing strengths in Environmental Data Science at TCUs.

Faculty understand unique cultural and community challenges (food, childcare, travel time, etc.)
Providing education beyond the classroom (research experiences, hands-on or independent research courses)
Students and faculty are grounded in their culture
Faculty value new ways of learning and teaching

Faculty are willing to provide support outside the classroom
Inclusion of community in student-led research projects
Supporting student-led research projects
Offering certificate programs
Seeking culturally affirming funding sources
Assess student outcomes using cultural models
Providing pathways for students to do data science without strong math backgrounds
Valuing place-based learning
The connection to community and experience, particularly for TCU faculty who are Tribal members
Looking within Tribal homelands to target studies which may benefit the local environment
Strong focus on community priorities

Table 4. Summary of responses listing strengths in Environmental Data Science at HBCUs.

Interdisciplinary research opportunities for students
Informal training available outside of the classroom
Using technology to introduce students to science in different contexts
Faculty understanding students' unique challenges (e.g., family pressure)
Connecting current students to graduates working in data science fields
Funding sources that support real world interdisciplinary research
Teaching data science across a breadth of fields/disciplines
Knowledge that cultural relevance and context might change when the research institution grows (i.e., R1)
Seeing a shift from citizen science to community science
Centering curricula in social justice and inclusion
Computational thinking
Partnerships with other HBCUs and MSIs, non-profits, and commercial industry
Investments in institutionally facilitated or operated cyberinfrastructure for training and research in data science

OPPORTUNITIES: What could we be doing more of to support students?

Considering these strengths, there continue to be opportunities that emerge to further support institutions in this work. Tables 5 and 6 summarize the opportunities brought forward by working group members and form respondents by MSI type.

Table 5. Summary of responses listing opportunities in Environmental Data Science at TCUs.

Gaining buy-in from administration and faculty to build or improve curriculum
Making student experiences more culturally relevant so students can see themselves as data scientists
Building culturally aligned career pathways for students to serve community needs

Leveraging existing resources to support the management of big data sets and coding
Demonstrating potential career opportunities in data science
Providing scaffolded or layered teaching approaches to support students in independent projects
Creating a curriculum repository across TCUs that matches students at different skill levels
Building connections through alumni networks, academic journals, etc.
Opportunities to go beyond collaboration in multi-million-dollar research projects (i.e., PI, or other)
Seeking culturally affirming funding sources
Transitioning Indigenous data sovereignty away from federal agency management
Partnering with more federal agencies and national organizations
Opportunities to also support TCUs in the capacity to handle grants/research
Providing skill building opportunities for the faculty who instruct students

Table 6. Summary of responses listing opportunities in Environmental Data Science at HBCUs.

Creating opportunities for early exposure to data science core concepts regardless of math background
Developing natural leaders in data science (e.g., clubs)
Supporting data hubs across HBCUs
Making existing resources more known with information about who to contact
Integrating fieldwork that aligns with classroom and workforce
Focus on recruiting more students in data science (marketing the field)
Building connections through alumni networks, academic journals, etc.
Providing scaffolded or layered teaching approaches to support students in independent projects
Connecting data science and programming (curriculum is sometimes old)
Gaining buy-in from administration and faculty to build or improve curriculum
Reducing the gap in public health institutional data
Leveraging the National Student Data Corps (NSDC)
ORCID for all faculty and anyone publishing a paper; publishing all academic work with a DOI
Expert Mentoring from the workforce of those who are from diverse ethnicities
Gaining buy-in from administration and faculty to build or improve cyberinfrastructure that enables real-world, large-scale data science skill development
Exposing students to the private sector

ASPIRATIONS: What do we aspire to do (or become) in the future?

By considering existing strengths and available opportunities through institutions and partner organizations, project participants were able to envision a future that builds on collective synergies. Tables 7 and 8 summarize the aspirations for future environmental data science activities provided by working group members and form respondents by MSI type.

Table 7. Summary of responses listing aspirations in Environmental Data Science at TCUs.

Sharing data in culturally appropriate and ethical ways
Indigenous data sovereignty
Community and self wellness (e.g., cultural immersion/celebration, nation-building, food sovereignty, stable jobs)
Using Indigenous values/frameworks to support students in courses as well as research
Seeing the students as scholars (not just as “research assistants” but also colleagues)
Increasing the use of real world examples and contexts that are centered on student interests and goals
Students understanding how data can/will be applied to the real world
Having Tribal data access approved by tribes
Funding to integrate language and culture with science so that community and students learn from each other
Additional funding to support students, infrastructure, and supplies
There being an entity that represents all Indigenous peoples regarding housing data
Collaborating to develop common metrics to work toward
Developing a stronger network between TCUs and partners, one in which knowledge and data have specific pathways for access

Table 8. Summary of responses listing aspirations in Environmental Data Science at HBCUs.

Flexible curriculum that can respond to real world changes (not just classroom decontextualized contexts)
Having more institutional support (e.g., time release, modules for curriculum)
Identifying tools and skills that students need (e.g., R, Python)
Faculty salaries are balanced with a living wage
There is trust that data are used "for the people" and/or "for social good"
Students have the data analysis skills to make real world changes
Having bridge programs/pipelines for students to go to graduate school
Seeing the students as scholars (not just as “research assistants” but also colleagues)
Integration of community at all levels (institution, faculty, students)
Funders increasingly understand that changes take time
Preparing students before entering college
Having access to a unique data set to succeed in R1 funding applications
Doing outreach to encourage interest and success in data science career opportunities
Increasing collaborations with better funded institutions
Having full scholarship funding
Ensuring that students can financially support themselves while studying

RESULTS: How will we know we got there?

As these institutions harness their strengths and ongoing opportunities, how will they know if they were able to reach their aspirations? How will success be defined? Tables 9 and 10 list the results provided by working group members and form respondents by MSI type.

Table 9. Summary of responses listing results from Environmental Data Science programs and activities at TCUs.

TCUs have inclusive, Indigenous-based course curriculum
Students have access to jobs and careers that they dream of
Students are increasingly pursuing and attaining advanced degrees in the field
Students know how to apply what they are learning/learned to their communities
Students can contextualize what they are learning in the classroom
Students have knowledge and understanding of their content area
Students know how to apply what they are learning/learned to their communities
Students are good problem solvers
Students are good computational thinkers and data analysts
Students are good communicators
Students are good collaborators
Students are well-rounded (aware of social, economic, cultural and political issues)
Students have the confidence to share knowledge with their peers/community members
Students have confidence in their ability to do research and work with data
Students can pursue experiences that might have been intimidating prior to going to college
Students accomplish more than they knew they were capable of
Students feel empowered to create new knowledge and solutions for their community
Students are in good health
Students experience spiritual growth
Students are living a good life
Opportunities tailor to student needs (not the other way around)
Data gathered is useful to the community
Community is happy
Community and Tribal Councils are given the opportunity to hear about research
Indigenous communities are no longer seen as “areas of expertise”, but “experts”
Students can bring their community knowledge to the research process
Students are giving back to our communities
Lifelong learning: students coming back to the lab

Table 10. Summary of responses listing results from Environmental Data Science programs and activities at HBCUs.

Success is defined as accessible to all

Diverse admissions
High percentage of students attending graduate school/professional school and completing successfully
Students are satisfactorily moving along their pathways
Students completing degree attainment
Students are finding and securing a thriving career
Students leave motivated, prepared, and more sought-after for their future education (e.g. graduate school) and career aspirations
Good career after graduation
Students know how to find and use resources to achieve their goals
Students have the skillset to reach their aspirations
Students gain knowledge and understanding of the subject matter
Demonstrate analytical skills in using data to solve public policy problems
Students understand their abilities and those of others
Students can solve real world problems
Students are working with cutting edge and current technology and intentional data sets
Students not only understand the methods, but also the business, social, and political context within which decisions that rely on the data analysis are made.
Students can take any scenario or use-case and figure out the approach effectively and efficiently.
Students understand their own needs and those of others
Students have an open mindset
Students see challenges and setbacks as new opportunities
Students know their learning capabilities.
Students can accurately identify their aspirations
Students are envisioning and embarking on bold, ambitious pathways
Students are self-directed
Students develop and leverage a passion for using data for positive societal, climate, community, and environmental impact.
Students like the faculty
Asking students what they want/need and how to define success

Common Themes across both Groups

Some themes emerged across both groups when the sentiments expressed were analyzed for full or partial alignment. Integrating community in Environmental Data Science (EDS) research was a common aspiration across TCUs and HBCUs. For TCU WG members, additional funding to integrate language and culture with science so that the community and students learn from each other was a recurring theme. For HBCU WG members, the aspiration to integrate the community reflected who should be involved, and included at all levels (institutional, faculty, and student). In addition, HBCU WG members echoed the need for additional and more flexible funding to support their efforts and the needs of their students.

There was also partial alignment for a common theme related to mathematics. Both WGs expressed the need for EDS to be more flexible in the math requirements to prevent these requirements from becoming a barrier to successfully navigating a degree in EDS, as it can be for other STEM fields. For example, there was interest in creating early pathways in data analytics which emphasize data gathering, cleaning, analysis, and visualization. At the same time, data science programs could work with mathematics departments to customize a shorter prerequisite-free mathematics sequence that includes the linear algebra and calculus skills needed for understanding dimensionality reduction and machine learning. For TCUs, providing pathways for students to do data science without strong math backgrounds was identified as an existing strength. This implies that some TCUs are already doing this. For HBCUs, creating opportunities for early exposure to data science core concepts regardless of math background was reported as an opportunity, or something they could be doing a little more of.

Priorities Moving Forward across both institutions

Some fully aligned for both Opportunities and Aspirations conveyed in the SOAR inquiry rose as "most important" items across both groups (TCUs and HBCUs) are proposed here as possible priorities moving forward. These include:

- Gaining administrative support and buy-in from administration and faculty to build or improve curriculum;
- Building connections outside the institution through alumni networks, academic journals, etc.;
- Having scaffolded or layered pedagogy to support students in independent projects; and
- Aspire to see the students as scholars, not just as "research assistants, but also as colleagues."

Opportunities for Future Investment

Each institutional group had its own key priorities, as well as its own values around how the work should be done. The recommendations listed here should represent possibilities for future investments in programs that support these institution types. For example, a large investment has been made into HBCUs, both private and National Data Science Alliance funding, to develop and share data science programming, research, and curricula. The possibility of investing in this level of coordination for TCUs was identified as an important future consideration. However, a key step toward this goal is to build relationships with tribes, especially around data governance.

The overlap in themes identified here could be used to hone in on opportunities that make strides in areas benefitting multiple communities. For example, investing in research and innovation through community-research partnerships in data science could benefit both TCUs and HBCUs or promote minority success at other institutions. Furthermore, the overlap in priorities could represent opportunities to invest in broader collaboration across contexts, for example, sharing practices for scale and sustainability, creating and leveraging networks of support across administration, alumni, and other sectors.

Next Steps & Invitation to Contribute

The project leadership team has taken the findings from this project and the resulting SOAR analysis to identify a number of next steps. These include:

- Raise awareness of activities being undertaken by organizations that can help advance shared aspirations
- Support the development of proposal submissions led by HBCUs and TCUs in the areas identified as potential opportunities
- Provide opportunities for others to add their comments and ideas to these recommendations

This report is meant to serve as a living document where members of the broader community can comment and add their expertise and perspectives. We invite you to add yours. After 30 June 2023, members of the project leadership team closed the report to further comment and transitioned the document to the ADSA Zenodo Community. The report link will be provided on our project team websites. This Google doc will re-open for public comment on 1 September 2023. We will revisit and incorporate any new content from the Google doc into the Zenodo document every six months until December 2024.

Works Cited

AlShebli, B.K., T. Rahwan, W.L. Woon. 2018. The preeminence of ethnic diversity in scientific collaborations. Nature Communications 9(1): 5163. https://doi.org/10.1038/s41467-018-07634-8

Barnes, A., A. Luh, M. Gobin. 2021. Mapping Environmental Justice in the Biden-Harris Administration. Center for American Progress https://www.americanprogress.org/issues/green/reports/2021/02/04/495397/mapping-environmental-justice-biden-harris-administration/.

Benjamin, R. 2019. Race after technology: Abolitionist tools for the new jim code. John Wiley & Sons.

Carroll, S.R., I. Garba, O.L. Figueroa-Rodriguez, J. Holbrook, R. Lovett, S. Materechera, M. Parsons, K. Raseroka, D. Rodriguez-Lonebear, R. Rowe, R. Sara, J.D. Walker, J. Anderson, M. Hudson. The CARE Principles of Indigenous Data Governance. Data Science Journal 19(1): 43. https://doi.org/10.5334/dsj-2020-043

Finley, K. 2017. Diversity in Open Source is Even Worse than in Tech Overall. Wired. Available: https://www.wired.com/2017/06/diversity-open-source-even-worse-tech-overall/

Hampton, S.E., C.A. Strasser, J.J. Tewksbury, W.K. Gram, A.E. Budden, A.L. Batcheller, C.S. Duke, J.H. Porter. 2013. Big data and the future of ecology. Frontiers in Ecology and the Environment 11(3): 156-162. https://doi.org/10.1890/120103

Hong, L. and S.E. Page. 2004. Groups of diverse problem solvers can outperform groups of high-ability problem solvers. Proceedings of the National Academy of Sciences 101(46): 16385-16389. https://doi.org/10.1073/pnas.0403723101.

National Academies of Sciences, Engineering, and Medicine (NASEM). 2019. Minority Serving Institutions: America's Underutilized Resource for Strengthening the STEM Workforce. Washington, DC: The National Academies Press. https://doi.org/10.17226/25257.

National Center for Science and Engineering Statistics. 2019. Women, Minorities, and Persons with Disabilities in Science and Engineering: 2019. Special Report NSF 21-321. National Science Foundation, Arlington, VA. Available: www.nsf.gov/statistics/wmpd/

Noren, L., G. Helfrich, and S. Yeo. 2019. Who's building our AI? Research Brief. Available at: https://www.obsidiansecurity.com/whos-building-your-ai-research-brief/

Rawlings-Goss, R., (Boots) Cassel, L., Cragin, M., Cramer, C., Dingle, A., Friday-Stroud, S., Herron, A., Horton, N., Inniss, T. R., Jordan, K., Ordóñez, P., Rudis, M., Rwebangira, R., Schmitt, K., Smith, D., and Stephens, S. 2018. Keeping Data Science Broad: Negotiating the Digital and Data Divide Among Higher Education Institutions. Mathematics and Statistics Faculty Publications. 64. https://scholar.valpo.edu/math_stat_fac_pubs/64

Silbert, J. H., & Silbert, T. (2007). SOARing from SWOT: Four lessons every strategic plan must know. *Innovation Partners International*.

Starck, J.G., S. Sinclair, J.N. Shelton. 2021. How university diversity rationales inform student preferences and outcomes. Proceedings of the National Academy of Sciences. 118(16): e2013833118. https://doi.org/10.1073/pnas.2013833118

Tilghman, S., B. Alberts, D. Colon-Ramos, K. Dzirasa, J. Kimble, H. Varmus. 2021. Concrete steps to diversity the scientific workforce. Science 372(6538): 133-135. https://doi.org/10.1126/science.abf9679

Xia, J., J. Wang, S. Niu. 2020. Research challenges and opportunities for using big data in global change biology. Global Change Biology 26(11): 6040-6061. https://doi.org/10.1111/gcb.15317.

Appendices

Appendix A: Project Justification

It is now broadly recognized that women and U.S. racial and ethnic minority groups are acutely underrepresented in STEM fields (National Center for Science and Engineering Statistics 2019). While a lack of diversity has been shown to limit innovation (AlShebli et al. 2018, Hong & Page 2004), this evidence has overshadowed the need to work towards diversity and inclusion as a moral responsibility to achieve justice in these fields (Starck et al. 2021). There is a growing need to remove barriers to careers in STEM in order to increase not just the impact of science on society, but to ensure its benefit for all (Tilghman et al. 2021). Within the data sciences, it was believed that free, open access data and products would foster diversity, but a GitHub study found that 95% of its users were male and only 16% identified as minorities in the country where they live (Finley 2017). More recently, a study found the representation of women and Black colleagues in data science research institutes well below the U.S. population and U.S. faculty overall, with only 26% and 1% representation, respectively (Noren et al. 2019). Because of the nature of big data, collective action is urgent and imperative to work alongside the marginalized faculty and student communities who are disproportionately impacted by its use, while finding ways to empower those communities to advocate for their own data (Carroll et al 2020, Benjamin 2019).

Multiple barriers, including the affordability of and access to new technologies and the lack of educational opportunities for diverse students, have led to a digital divide that demonstrates the inequities associated with who have access to data and the skills to extract knowledge from those data (Rawlings-Goss 2018). Data science education at Minority Serving Institutions (MSIs) can narrow this divide by ensuring a diverse student population has the knowledge and skills required to address issues of social and cultural relevance. MSIs have proven to be effective at recruiting, retaining, and preparing underrepresented students for the STEM workforce when compared to predominantly white institutions. Historically Black Colleges and Universities (HBCUs), which comprise only 3% of postsecondary institutions and enroll only 9% of Black undergraduate students, graduate the highest number of Black students who go on to earn STEM doctorates (NASEM 2019). Tribal Colleges and Universities (TCUs) have integrated American Indian culture, languages, and traditions into their learning environment to improve student success, with 86% of students completing their degree compared to <10% who attend other institutions (NASEM 2019).

As data science education and leadership organizations, the Academic Data Science Alliance (ADSA) and the Environmental Data Science Inclusion Network (EDSIN) brought together representatives from HBCUs and TCUs to empower and support equitable data science education and co-create solutions that ameliorate the digital divide. Our vision is to ensure all students have the resources and skills required to be effective in applying data science throughout their careers, ensuring their competitiveness in the workplace where they will be active participants in using their knowledge and skills to influence policies and decisions that may directly impact their communities.

We focused specifically on the application of these skills to the environmental sciences as there has been significant growth in large ecological datasets and environmental synthesis projects over the last decade resulting from such projects as the National Ecological Observatory Network (NEON), the Long-Term Ecological Research Network (LTER), and the Environmental Data Initiative (EDI; Hampton et al. 2013, Xia et al. 2020). The National Science Foundation has also recently funded a new open environmental data synthesis center, Environmental Data Science Innovation & Inclusion Lab (ESIIL), with "an important focus on diversifying the workforce that has the skills and background to use open biological and other environmental datasets." Environmental and climate justice issues have come to the forefront in recent years with the promise of many new jobs in this sector proposed by the current Biden administration (Barnes et al. 2021). In addition, a focus on environmental data science allowed us to build on existing efforts of the Environmental Data Science Inclusion Network (EDSIN). EDSIN was initiated during a conference hosted in April 2019, with funding from the NSF INCLUDES program, where participants examined justice, equity, diversity, and inclusion at the intersection of the environmental and data science fields. During this event, attendees established the network to begin addressing priority areas that emerged from the conference discussions, including: accessibility of courses and relevant tools; full integration of data ethics; building and implementing culturally and socially relevant curricula; co-creation and collaboration for resource development; and demonstrating relevance of environmental data science to diverse communities (Crall et al. 2020)

APPENDIX B: PROJECT PARTICIPANTS

PARTNER ORGANIZATIONS

The Academic Data Science Alliance (ADSA) is a community of academic data science leaders, practitioners and educators who connect and share their data-intensive approaches and responsible applications. Current membership in ADSA exceeds 650 individuals across more than 50 colleges and universities. The opt-in Data Science Community Newsletter funded by ADSA reaches over 8,600 inboxes. Dr. Micaela S. Parker (PI and Leadership Committee Chair) is the Founder and Executive Director of ADSA. Prior to launching ADSA, Dr. Parker developed data science research and training programs as an Executive Director for the eScience Institute at the University of Washington and Program Coordinator for the Moore-Sloan Data Science Environments.

The Native BioData Consortium (NBDC) is the first Indigenous-led research center and biological-data repository (or "biobank") initiative driven by Indigenous scientists and community members within a tribal nation in the United States. It is proximally located to >13 TCUs. Dr. Krystal Tsosie (Co-PI and TCU WG Chair) is an Assistant Professor at Arizona State University with expertise in data science and genetic epidemiology and a former instructor at Turtle Mountain Community College. As an Indigenous (Diné/Navajo Nation) geneticist and bioethicist working with several tribal communities, she is a leader in building capacity for Indigenous genomics education, working as a co-facilitator for SING (Summer Internship for Indigenous Peoples in Genomics) and NBDC.

The Atlanta University Center (AUC) Data Science Initiative is a collaboration envisioned by the presidents of the AUC Consortium member institutions that are Historically Black Colleges and Universities (HBCUs): Clark Atlanta University, Morehouse College, Morehouse School of Medicine, and Spelman College. Leveraging the AUC's rich history in social justice, the Initiative aims to increase the number of Blacks with credentials in data science and develop innovations in data science that promote the ethical use of data to benefit society. Dr. Talitha Washington (Co-PI and HBCU WG Chair) is the inaugural Director of the Initiative and oversees and provides strategic direction and coordination of data science education, research, and industry engagement.

The Carpentries was formed to teach foundational computation and data science skills through short impactful workshops. All course materials are open source, in the public domain, and free to use and contribute to. Since 2012, The Carpentries have run 2,700 workshops in 71 countries and trained 2,800 volunteer instructors to deliver 45 collaboratively developed, open lessons to 66,000 novice learners at 98 member sites and beyond. Dr. Kari L. Jordan (Co-PI) is their Executive Director and leads efforts to measure the impact workshops have on learners, and develops metrics for equity, inclusion, and accessibility. Dr. Alycia Crall (Partners Support Lead) is the Director of Community at The Carpentries and was PI on the NSF INCLUDES grant that initiated EDSIN. She served as the network's community manager from 2019 through 2022.

The Institute for a Racially-just, Inclusive, and Open STEM Education (RIOS) works to amplify the value and reach of open education in STEM by helping organizations and projects align their resources and practices with the principles of anti-racism, equity, social justice, and inclusion. They see open education as an approach and mindset to transform teaching and learning to center the needs of underrepresented and marginalized learners and instructors who have been systematically excluded from the benefits of traditional educational systems. Dr. Carrie Diaz Eaton (Member, Leadership Committee) is the PI of the RIOS Institute, a co-organizer of the EDSIN conference and also serves as Chair of the Committee for Minority Participation in Mathematics for the Mathematical Association of America. Diaz Eaton also represented QUBES Hub as the former Partnership and Communications Director.

The National Ecological Observatory Network (NEON), a continental-scale observation facility and program funded by NSF, collects over 180 free and open access data products and samples designed to research issues at the forefront of the environment and society. NEON has offered workshops for MSI faculty and students interested in using NEON data, and it has been working to develop socially and culturally relevant curricula using NEON and other open environmental datasets. Dr. Claire Lunch (Member, Leadership Committee) leads training initiatives for NEON. She was the co-PI on the NSF INCLUDES grant that initiated EDSIN.

Center for Scientific Collaboration and Community Engagement (CSCCE) is a research and training center to support and study the emerging field of scientific community engagement - how to nurture successful STEM communities. CSCCE's Executive Director, Lou Woodley, has extensive experience building collaborative infrastructure, and designing and facilitating engaging, inclusive meetings - including focus groups, online webinars and workshops. CSCCE worked with the Leadership Committee to develop and facilitate a total of 5 half-day workshops for the HBCU and TCU WGs.

Sara Bolduc Planning and Evaluation LLC (SBPE), served as the external evaluation consultant for this project. SBPE is a private research firm located in Honolulu, Hawai'i, that offers policy and program evaluation, as well as strategic planning services in the Pacific Region and the greater United States. SBPE President Dr. Sara Bolduc has expertise in survey design and administration, and evaluating large, collaborative science efforts. The SBPE team (Sara Bolduc and Kendi Ho) worked with the Leadership Committee and CSCCE to develop a SOAR (Strengths, Opportunities, Aspirations, Results) analysis of the topics forefronted by the WG meetings; created and administered the form to gather broader community input for the final Recommendations Report framed by the SOAR model; and assisted in the creation of a post-workshop survey of WG participants.

WORKING GROUP MEMBERS

Dana Gehring

Oglala Lakota College

Faculty member at tribal colleges (currently Oglala Lakota College, previously Sinte Gleska University) for 16 years in total in biology and environmental science.

Kaitlyn Haskie

Diné College

Kaitlyn Haskie serves as a project manager, research advisor, and mentoring professional at Diné College, a tribal college on the Navajo Nation. In her capacity as a project manager for the grant, Undergraduate Readying for Burgeoning Research for American Indian Neuroscientists, URBrain, she has gained skills unique to mentoring at a minority serving

Jessica Hernandez

University of Washington Bothell

Jessica Hernandez completed a higher academic degree in the environmental sciences and currently serves as an instructor and postdoctoral research fellow at an MSI that is working on a program to serve a college within the Yakama nation but is not a TCU.

Guanyu Huang

Spelman College

Guanyu Huang is an assistant professor of Environmental and Health Sciences at Spelman College. His research interest includes incorporating data science into air quality, climate and remote sensing research. He is also interested in data science education.

Brandi Kamermans

Northwest Indian College

Brandi Kamermans is a Postdoc and Molecular Researcher at Northwest Indian College on the Lummi main campus at the Salish Sea Research Center (SSRC). She monitors harmful algal bloom species in Bellingham and Lummi Bay using molecular techniques. She is currently working on refining qPCR methods to detect and quantify Longfin Smelt (Hoolies) in the Nooksack River and Bellingham Bay.

Erdi Kara

Spelman College

Erdi Kara is an applied mathematician by training. He has experience in deep learning, particularly in vision problems, including image recognition, object detection and segmentation. He is fluent in Python, C++ and MATLAB.

George Middendorf

Howard University, Department of Biology

George Middendorf received his Ph.D. (Zoology, Tennessee-Knoxville) and completed a postdoc (Iowa), he joined the faculty at Howard University until retiring in 2021. He has published on the ecology study of Yarrow's spiny lizards, the defensive blood-squirting behaviors of horned lizards, herpetofauna of Maryland, West Virginia, Bolivia, Peru and Suriname, urban environmental issues, and urban sloths in Suriname. He has taught a wide variety of Biology courses, including a number of interdisciplinary courses.

Teresa L Newberry

Tohono O'odham Community College

Teresa Newberry is Chair of Science and Health at Tohono O'odham Community College (TOCC) where she has served as a Faculty member in Biology and Natural Resources since 2005. She is also Principal Investigator (PI) and Project Director on an NSF TCUP ICE-TI grant "Pathways to Indigenous STEM". Her academic background includes a Ph.D. in Biology from the University of New Mexico and an M.S. in Natural Resources from the University of Michigan and her research interests include physiological plant ecology, geochemistry, and climate-vegetation relationships.

Melinda Otto Neville

Leech Lake Tribal College

Melinda Otto Neville's research interests include environmental monitoring systems (EMS) data with spatiotemporal trends analyses. Current projects at Leech Lake Tribal College include remote EMS soil and climate data, local weather data, and macroinvertebrate bioassessments. Her lab's work on restoring fire in fire dependant plant communities supports community forest stewardship and a return to desired vegetative conditions within Leech Lake Nation.

Ruth Plenty Sweetgrass-She Kills

Nueta Hidatsa Sahnish College

Ruth Plenty Sweetgrass-She Kills earned her A.S. in Environmental Science from Fort Berthold Community College (Nueta Hidatsa Sahnish College). She has worked in the sciences as a staff member or faculty member at two TCUs (Nueta Hidatsa Sahnish College and Fort Peck Community College). Her research collaborations have included tribal programs, state and federal agencies, sister TCUs, and other institutions of higher education.

Joey Riley

Leech Lake Tribal College

Joey Riley is a natural science instructor at Leech Lake Tribal College. He grew up in northern Minnesota and has spent four summers conducting wildlife and forestry research here. His expertise is in applying statistical and spatial analyses to explore wildlife biology and forest ecology.

Jeticia Sistrunk

Children's Hospital of Philadelphia; previously Spelman College

Jeticia Sistrunk is currently a Bioinformatics Scientist at the Children's Hospital of Philadelphia. During this project, Sistrunk was an Assistant Professor in the Department of Biology at Spelman College. Their work uses both molecular and computational approaches to study food and waterborne infections, and microbial communities relevant to human health. Their interests include STEM outreach and developing innovative curriculum for integrating computing techniques in biology courses.

Belin Tsinnajinnie

Santa Fe Community College

Belin Tsinnajinnie (he/him) is Diné and Filipinx from Na' Neelzhiin, New Mexico and an Assistant Professor in the Mathematics Department at Santa Fe Community College. Belin received his PhD in Mathematics at the University of Arizona with a doctoral thesis focused on notions of mathematical identity in the context of Indigenous and Latinx students. Belin served for eight years as a full-time faculty at the Institute for American Indian Arts which included developing and teaching a curriculum in quantitative reasoning.

Krystal Tsosie

Native BioData Consortium

Krystal Tsosie (Diné/Navajo Nation), MPH, MA, PhD candidate is an Indigenous geneticist-bioethicist at Vanderbilt University, incoming faculty at Arizona State University. She co-founded the Native BioData Consortium, a nonprofit research institute for advancing Indigenous genomic and data sovereignty. She also co-founded IndigiData, an Indigenous data science workshop that utilizes environmental microbiome data in an Indigenous culturally-centered approach to include Indigenous data sovereignty and data ethics. She is a former STEM researcher and instructor at Turtle Mountain Community College.

Ethell Vereen

Morehouse College

Ethell Vereen, Jr. (he, him, his) is an Assistant Professor of Biology at Morehouse College with a STEM Education focus on diversity and inclusion in natural resources; and a research focus on environmental health, water quality, and environmental microbiology. Dr. Vereen was recently awarded a $300,000 Research Initiation Award from the National Science Foundation for his current project, "Metagenomic Approach to Assess Water Quality and Microbial Load Variability of an Urban Watershed," that has increased the research capacity at Morehouse College and provides additional training opportunities for the College's STEM students. He currently serves on several boards and foundations, including the board for the Environmental Leadership Program, NSF NEON Technical Working Group, and his exemplary contributions in the areas of undergraduate education, student learning and campus life have also been noted as he was awarded the 2019 Vulcan Teaching Excellence Award and recently recognized as one of 1000 Inspiring Black Scientists in America.

ClarLynda Williams-DeVane

Fisk University and North Carolina State Government

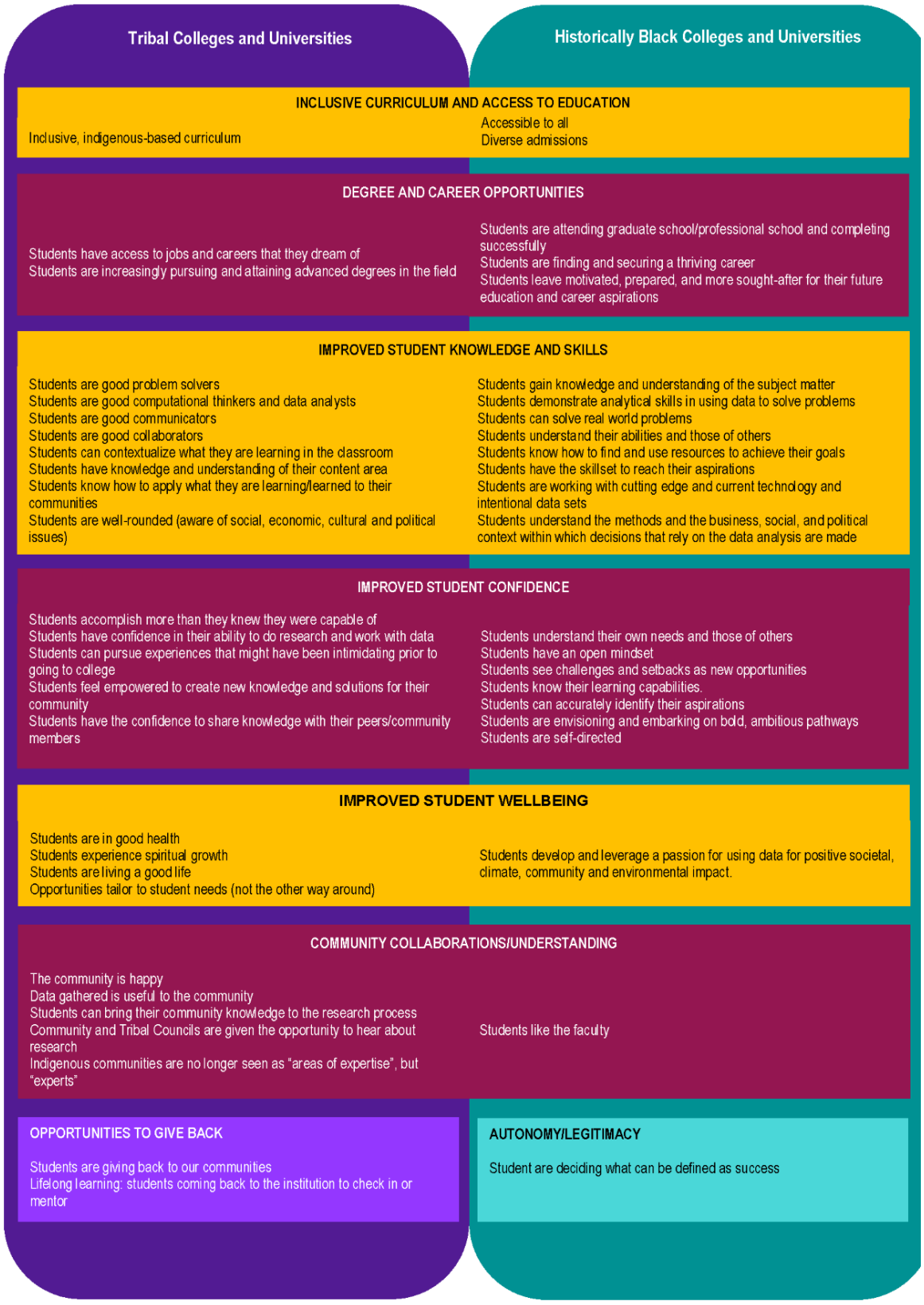
ClarLynda Williams-DeVane has served as a department chair in data science and now works in public health statistics and health equity. She is a health disparity informaticist in practice, but a bioinformatician by training. She has worked as an assistant/associate professor at NC Central University and as a Department chair of Data Science and Bioinformatics at Fisk University.

Talitha Washington

Clark Atlanta University & Atlanta University Center

Dr. Talitha Washington is the inaugural Director of the Atlanta University Center (AUC) Data Science Initiative, a Professor of Mathematics at Clark Atlanta University and an affiliate faculty at Morehouse College, Morehouse School of Medicine, and Spelman College. She is the Director and lead principal investigator of the NSF-funded National Data Science Alliance (NDSA) and the President of the Association for Women in Mathematics. She works across Historically Black Colleges and Universities (HBCUs) to increase the number of Blacks with expertise in data science and expand data science research that advocates for social justice. Washington is a former Program Director at the National Science Foundation (NSF) and a Fellow of the American Mathematical Society, the Association for Women in Mathematics, and the American Association for the Advancement of Science.

Appendix C: Assessing the Extent to Which Students at MSIs are Successful in Environmental Data Science



Appendix D: Educational Tools And Resources: Supporting Student Success In Data Science

Funding Opportunities
US Department of Agriculture (USDA)
National Institute of Food and Agriculture (NIFA)
National Science Foundation (NSF)
National Aeronautics and Space Administration (NASA)
AUC Data Science Initiative: seed grants support time for course development related to data science

Collaborators
Tribal Cultural Centers

Multiple resource types (e.g. data, curriculum, professional development, repository)
QUBESHUB for modular resources and community
SERC Science Education Resource Center at Carleton: https://serc.carleton.edu/index.html

Curriculum Resources
WestEd (institutional support and training)
Resource: man in the maze model for college's core values (Spider Conceptual Framework?)
Pedagogical resource: Universal Design https://udlguidelines.cast.org/
PHET University of Colorado Boulder - interactive simulations;
https://phet.colorado.edu/en/simulations/browse
Merlot http://www.hbcuals.org/findcoursematerials.html
http://www.hbcuals.org/cultural_collections.html
Ecological Society of America (ESA's 4DEE (4 Dimensional Ecology Education) Frameworks
https://www.esa.org/4DEE/

Professional Societies/Communities
Academic Data Science Alliance (ADSA) https://academicdatascience.org/
Environmental Data Science Inclusion Network (EDSIN)
Biological and Environmental Data Education (BEDE -new RCN-UBE network)
Environmental Synthesis Centers, e.g. NCEAS https://esiil.org/
RIOS Institute
BioQUEST

Open data sources
NEON https://www.neonscience.org/
Ocean Observatories Initiative (OOI) https://oceanobservatories.org/
National health databases (e.g. https://pearldiverinc.com/)
Environmental Data Initiative https://environmentaldatainitiative.org/
DataOne https://www.dataone.org/
Open Topography https://opentopography.org/
Federal data clearinghouse https://data.gov/

Professional development
The Carpentries: https://carpentries.org/
Example of workshop: https://www.accelevents.com/e/datadiaspora

Other Resources
Academic Journals: tribal college journal; https://www.jbhe.com/hbcus/
Social networking : https://hbcuconnect.com
Morehouse innovation and entrepreneurship center (https://mcecenter.com/)
protocols.io "A secure platform for developing and sharing reproducible methods."
Open-source tools for large datasets: preparation and cleaning, e.g. Dryad: https://datadryad.org/stash