Data Science for High School Computer Science: Identifying Needs, Gaps and Resources

While there has been intense interest in bringing computer science to all learners through such initiatives as Computer Science for All, the ability to solve complex real world problems through the application of computational skills is through data science. Yet data literacy skills are largely absent from educational systems, and while there have been attempts to introduce data science skills into computer science curricula, these attempts have lacked access to the knowledge, resources, training and accessible tools and data sources needed to develop and scale data literacy along with computer science literacy. Columbia University Data Science Institute / Northeast Big Data Innovation Hub, the New York Hall of Science and University of California San Diego propose to bring learning science experts, data science domain experts and tool developers together with education and private sector stakeholders to propose a set of priority implementation strategies intended to address the data science-computer science education gap. While the need exists throughout PreK-20 education, this effort will focus on the needs of high school students and teachers, as they are on the front lines of the rapidly changing workforce. The overarching purpose for this capacity building workshop will be to articulate a pathway toward data literacy and emphasize the unique and genuinely new dimensions of learning afforded by data science and how they create new opportunities for applying computational thinking, programming and habits of mind to new problems, learning and insights in STEM domains.

The 2-day, in-person workshop will include invited stakeholders from industry, state/local government, K-20 curriculum developers, instructors, students, and specialists in learning sciences and informal learning. The workshop will include four elements: 1) identifying the gaps of data science in computer science; 2) a series of presentations from tool and software developers and learning researchers to describe validated practices; 3) a collaborative brainstorming session to draft a fundamental set of essential concepts; and 4) authoring of a white paper to articulate the path forward. Deliverables will be made available on the Northeast Big Data Innovation Hub data literacy website to disseminate products widely. These working documents will form the basis for deployment recommendations and trajectory for infusing data science into computer science at the high school level and ultimately throughout formal education in the United States.