

HDR DSC: Collaborative Research: *Connecting the Dots*

Development of a data science educational program to increase the data acumen of undergraduate engineering majors

PI: Jeffrey R. Errington; Co-PIs: Liesl Folks, Kristen Moore, Bina Ramamurthy, and Erin Rowley



Introduction

There is significant demand for a workforce that is proficient in data science and analytics. Employers seek graduates with an ability to (1) understand, interpret, and analyze data, (2) effectively communicate results that stem from the analysis of data, (3) practice the ethical use of data, and (4) apply data science concepts to solve practical problems with real-world relevance. While the dissemination of data science competencies has been emphasized in some disciplines (e.g., computer science), the broad delivery of these skills to college graduates has been slow to evolve. The aim of this project is to develop and implement a scalable, innovative program, termed "Connecting the Dots", for delivery of data science competencies to students pursuing an undergraduate engineering degree.

Connecting the Dots (CTD) is a highly collaborative project between the flagship schools in the SUNY and CCNY systems. CTD teams the University at Buffalo (UB) with the City College of New York (CCNY). The signature academic data science track created by the CTD project team is an undergraduate micro-credential (digital badge) program that is readily integrated with any engineering major at the undergraduate level. A broad range of community partners are served via novel Data Science Community Labs, which act as "pop-up" summer facilities on the UB and CCNY campuses wherein students perform internship projects for community partners who have challenging data science problems for students to work on but are not well-positioned to host a conventional intern. The team's ultimate scaling objective is to develop a program that is easily adopted by other SUNY and CCNY campuses that host 4-year engineering programs and by campuses outside of New York State with similar degree program structures.

Educational Program

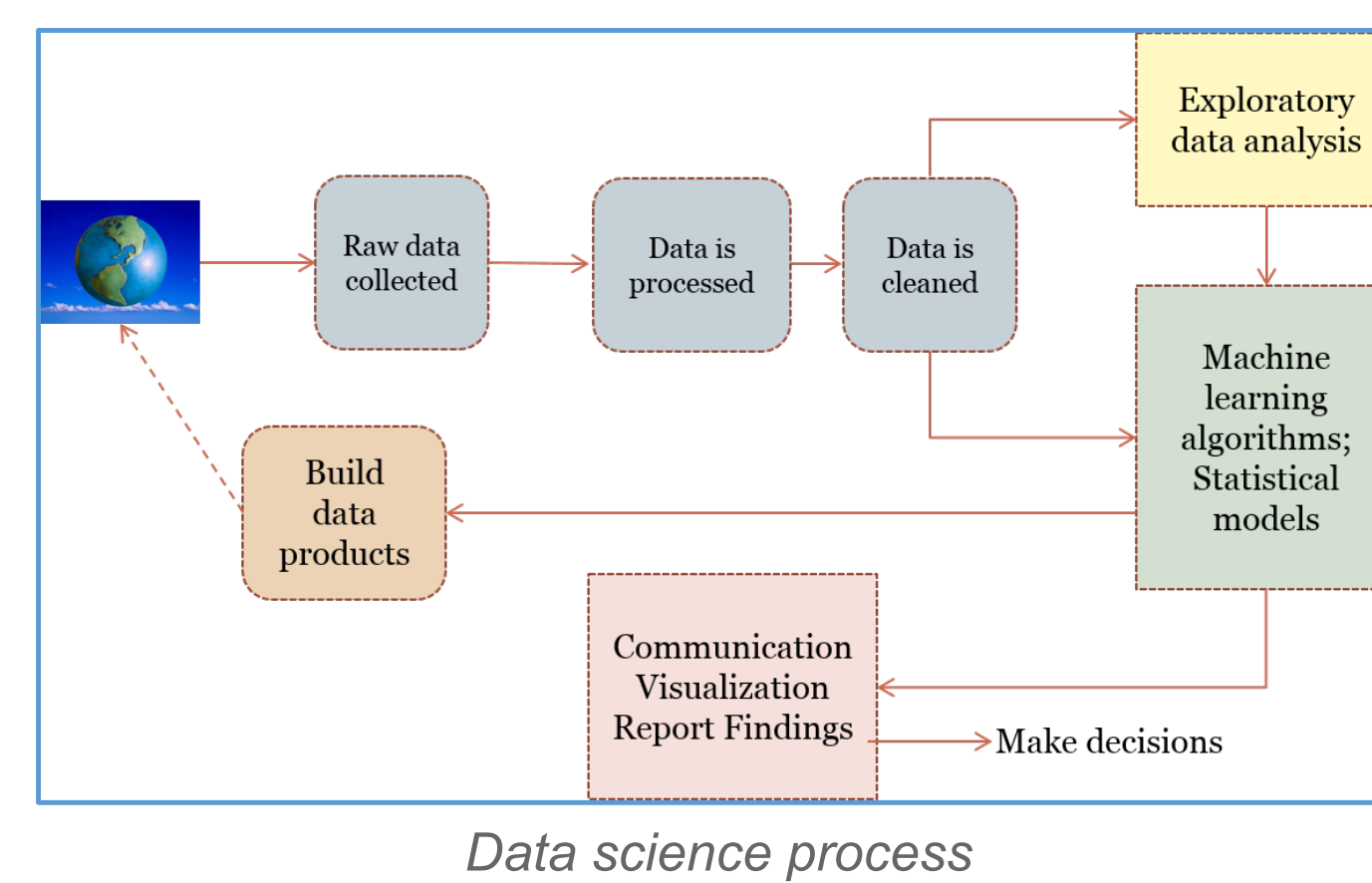
The digital badge program includes four key elements:

1. Course 1: Introduction to Data Science
2. Course 2: Communicating with Data
3. Data Science Internship
4. Capstone Project

Course 1: Introduction to Data Science

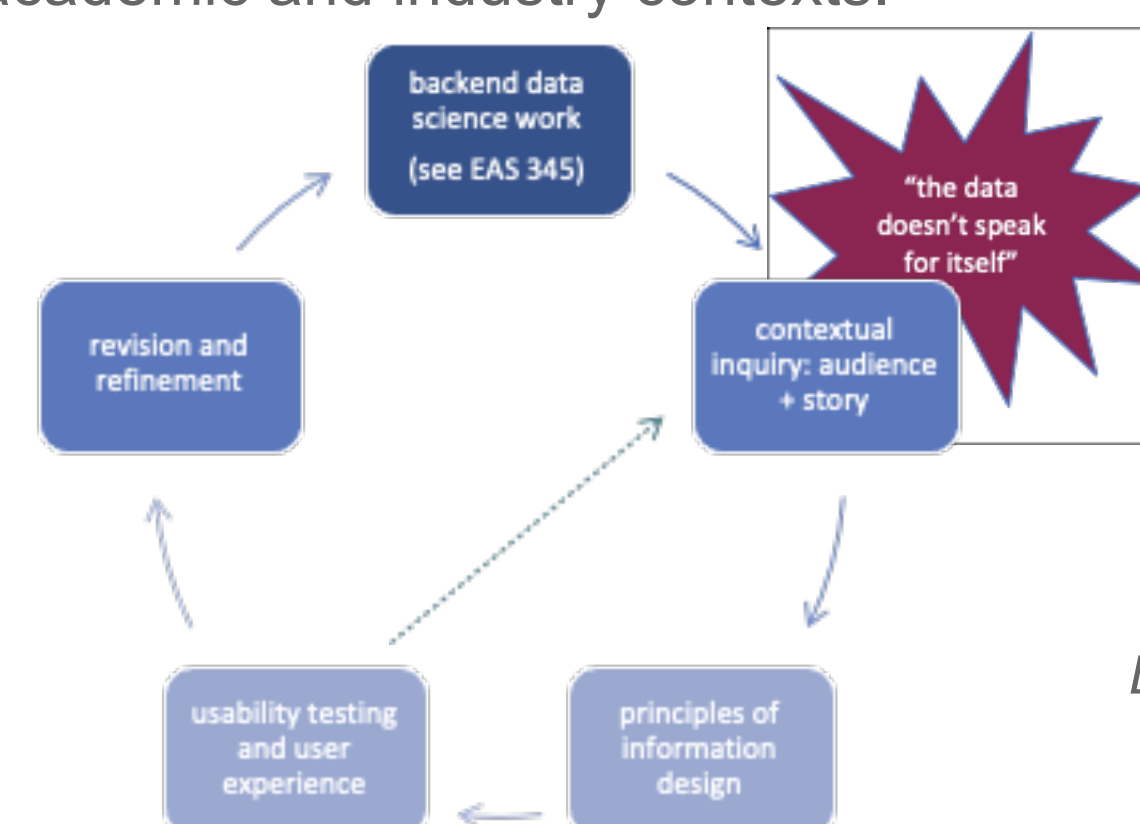
This course is an introductory level course in data science, with an emphasis on fundamental concepts in data science. The course focuses on three major themes: data characteristics, data science pipeline, and data-driven applications. Topics include: Data diversity, data products, data collection methods, data cleaning, and formatting, storing and sharing data, privacy, and confidentiality of data, data security, small data analysis, and statistical analysis. All concepts covered are illustrated via hands-on projects and problem-based learning activities, analyzing real-world data sets to develop data science skills. (3 credits)

Prerequisites: Calculus 2 and programming course



Course 2: Communicating with Data

This course introduces students to the principles of data communication with a focus on how understanding and practicing visual communication aids in our ability to communicate complex information. Working both graphically and linguistically, students refine their writing and presentation skills for professional contexts through the iterative design of visuals, including data displays, slide decks and power points, and written genres like journal articles and white papers. This project-based course engages students in application and practice that directly applies to their experience as professionals in both academic and industry contexts.



Data Science Internship

Students are required to complete a 10-week full-time summer (or equivalent) internship wherein they apply data science concepts to solve practical problems with real-world relevance. We foster two internship modes.

Conventional Internships: students work on site at the work site of the company that sponsors their internship, such that they experience the workplace first-hand and absorb the professional culture of the organization and sector.

Data Science Community Labs: "pop-up" summer facility on UB campus wherein students perform internship projects for community partners who have challenging data science problems for students to work on, but are not well-positioned to host a conventional intern (e.g., small non-government organizations (NGOs), smaller state or local government entities (LGs), and small or medium enterprises (SMEs)).



Examples of Community Partners with Internship Projects

Capstone Project

The capstone activity has students develop and produce a short video or equivalent multimedia module (e.g., a Jupyter notebook or a web page) that

- illustrates some of the work done,
- focuses on the relationship between experiences from the internship and key concepts in data science,
- shows appropriate use of data communication and visualization tools, and
- demonstrates how one or more professional competencies (teamwork, entrepreneurship, critical thinking, communications, collaboration, creativity, and ethics) was important to their internship work.



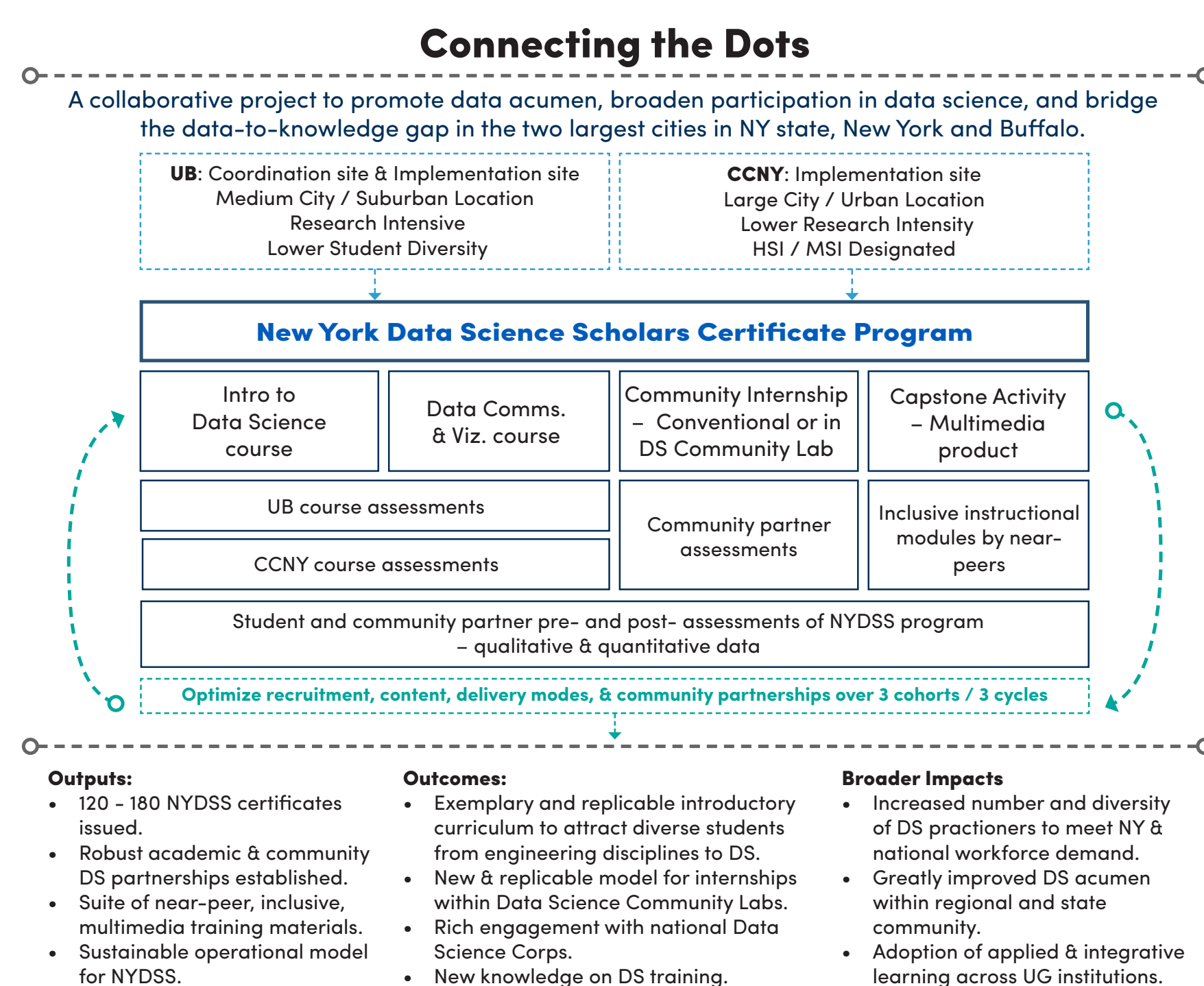
Recording Studio in the UB Silverman Library

Accomplishments

- Development of the course EAS 345 *Introduction to Data Science*. The course has been offered fall 2020, spring 2021, fall 2021, and fall 2022.
- Development of the course EAS 346 *Communicating with Data*. The course has been offered spring 2021 and spring 2022.
- Sought and obtained approval for the digital badge program. Such programs are referred to as micro-credentials at the University at Buffalo.
- Developed instrumentation and collected data, including a) a pre- and post-test survey, b) reflection and assessment instrument for the internships, and c) assessment guidelines for the capstone course.
- Developed and obtained approval for an IRB proposal regarding solicitation of feedback from students.
- Creation of a student-facing website outlining the benefits of the digital badge program.
- Guided two cohorts of students through the internship program. Sites included Moog, M&T Bank, People Inc, and Helix Intel.
- Engaged in productive discussions with CCNY partners.

Looking Forward

- Recruit additional students to participate in the Data Science digital badge program. Recruitment proved difficult during the pandemic while many students disengaged from experiential learning programs.
- Work with UB's Career Design Center to identify organizations with data science internship opportunities.
- Recruit additional organizations to leverage the Data Science Community Labs.
- Engage in program evaluation, including the recruitment strategy; participant interest, self-efficacy and identity; value of integrated Conventional / Community Lab internships; value of participant's Capstone Activity; value delivered to community organizations.



Schematic of project structure for Connecting the Dots