Collaboratory at Columbia: An Aspen Grove of Data Science Education

Isabelle A. Zaugg^{1,2} Patricia J. Culligan^{1,3} Richard Witten⁴ Tian Zheng⁵

¹Data Science Institute, Columbia University, New York City, New York, United States of America,

²Institute for Comparative Literature and Society, Columbia University, New York, New York, United States of America,

³College of Engineering, University of Notre Dame, Notre Dame, Indiana, United States of America,

⁴Columbia Entrepreneurship, Innovation, and Design, Columbia University, New York City, New York, United States of America,

⁵Data Science Education Working Group, Data Science Institute, Columbia University, New York City, New York, United States of America

Published on: Oct 28, 2021

DOI: <https://doi.org/10.1162/99608f92.53c4a1b4>

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

ABSTRACT

Many universities recognize the rapidly growing impact of data science in all fields of study and the professions and seek to embed this expertise widely across their educational offerings. There is often broad interest in developing new data science curricula, with some universities even allocating funds toward this purpose. Yet, it is often unclear what resources are needed for effective data science education, and how resources ought to be prioritized. Although university leadership might be aware of a growing number of successful data science acceleration programs and pedagogical models, many of which are either general purpose or specific to a particular discipline, there remains a lack of clarity about how these models might address their own specific needs. This article presents the *Collaboratory* Program at Columbia University (termed “the Collaboratory”), which is both a set of “data science in context” educational approaches, as well as a meta-model for an accelerator program that allows different institutions to respond flexibly to their own disciplinary heterogeneity in terms of data science educational needs. The novelty of the Collaboratory lies in its crowd-sourcing approach to creating new data science pedagogy and its ability to kindle transdisciplinary collaboration in doing so. By offering seed funding, it fosters proactive efforts to embed data science “in context” into more traditional domains through a cohort of compelling, transdisciplinary, crowd-sourced data science education proposals each year. Collaboratory educational offerings are required to be developed through a partnership between two faculty members, a data scientist and a domain expert from another field, or a larger team with complementary expertise. Over the past 5 years, the Collaboratory has supported the development of a wide spectrum of data science pedagogical models spread across more than 40 academic departments, centers, institutes, and professional schools at Columbia University. As a result, the Collaboratory has to date served the learning needs of more than 4,000 students. Furthermore, it has cultivated a thriving ecosystem that includes a funding mechanism and a community-support structure that all contribute to its agility and success. Here, we offer our experience and best practices in developing and managing the Collaboratory, which, we hope, will contribute to a blueprint for data science education leaders everywhere.

Keywords: data science education, transdisciplinary education, data science competency, data science literacy

1. Introduction

Data-intensive applications increasingly impact every domain of our lives, from developing better health care outcomes, to filling out a stronger roster on our favorite sports team, to building marketing campaigns that target niche audiences. In turn, it is clear that universities that wish to stay current must equip their students, across almost all domains, with relevant data science expertise (King, 2011; Lazowska, 2018; Moore-Sloan Data Science Environments: New York University, UC Berkeley, and the University of Washington, 2018; National Academies of Sciences, Engineering, and Medicine [NASEM], 2018; Van Dusen et al., 2019; Wing et al., 2018; Wing & Banks, 2019). We have also become more aware of the grave consequences of treating ethics

and social consequences as an afterthought, or an appendage, to data-driven services (Wing, 2018; Wing et al., 2018; Wing & Banks, 2019; Zook et al., 2017). This fuels a growing recognition that universities must ensure that data science education informs and is informed by the humanities, social sciences, and other traditional domains (Chun & Rhody, 2014; Moore-Sloan Data Science Environments, 2018; NASEM, 2018; Pawlicka, 2017; Wing et al., 2018; Wing & Banks, 2019). Furthermore, negative and inequitable consequences of data-driven innovation can, in part, be linked to the lack of diversity within the data science workforce, both in terms of demographics and disciplinary perspective. Universities are therefore being called on to expand data science education to a larger and more diverse population of learners (Lue, 2019; Moore-Sloan Data Science Environments, 2018; NASEM, 2018; Rawlings-Goss et al., 2018; West et al., 2019).

So how can universities best approach the challenges outlined here considering their own distinct faculty, student populations, schools and departments, resources, and educational philosophies? A number of approaches have been piloted and shared (Baumer, 2015; Cleveland, 2001; Hardin et al., 2015; NASEM, 2018; Yan & Davis, 2019; Zook et al., 2017). One celebrated example is University of California, Berkeley's Data 8: The Foundations of Data Science course, which offers one possible model for *undergraduate data science education*. Data 8 is notable in its rapid roll-out, benefiting from a university-level task force (Data Sciences Education Rapid Action Team, 2015). It was designed to be effective in reaching a large undergraduate population through a *standardized* format and large class sizes (Moore-Sloan Data Science Environments, 2018; Van Dusen et al., 2019). Its focus on students who have *not* previously taken courses in computer science or statistics has been key to its success. Serving as a single point of entry into data science skill sets, it is further complemented by 'connector courses' that bridge data science skills into different domains (Data Sciences Education Rapid Action Team, 2015; Van Dusen et al., 2019).

While approaches such as Data 8 have important merits, in this article we share *another* model for the rapid development of data science education programs, known as the *Collaboratory* Program at Columbia University (or "the Collaboratory," for short). Co-founded by Columbia Entrepreneurship and Columbia's Data Science Institute in 2016, the Collaboratory "supports the development of innovative, interdisciplinary curricula that embeds data or computational science into more traditional domains or the reverse, embeds business, policy, cultural, and ethical topics into the context of a data or computer science curriculum" (Collaboratory at Columbia University, n.d.). A short video introduction to the Collaboratory can be viewed [here](#). The Collaboratory crowd-sources innovative new course designs by providing seed funding to support transdisciplinary collaboration between data scientists and domain experts, called "Collaboratory Fellows." The Collaboratory jumpstarts data-driven pedagogy that otherwise would take years to develop.

Sixty-seven Collaboratory Fellows have launched 21 Collaboratory projects to date. These Fellows span 14 schools, 14 institutes and centers, and 26 departments at Columbia University, and the courses they have designed are impacting students from across Columbia's many programs. As of Fall 2020, the Collaboratory had supported the development of 36 new courses, out of which 26 have been taught one or more times. The

Collaboratory has also supported the revision of three existing courses, the development of two bootcamp/workshop series, five capstone projects, a multiyear lecture series, and a master's program in Environmental Health Data Science, among other outputs. A number of Collaboratory courses have been integrated into the core curricula in their field. Collaboratory course sizes range from six to 99 students. More than 4,000 students have taken Collaboratory courses as of the Fall 2020 semester. The Collaboratory's broad reach has also increased awareness of data science among university leadership. Collaboratory proposals require a commitment from department chairs and deans to continue offering a course after seed funds are spent. As such, a Collaboratory proposal creates an opening for faculty and university leadership to pursue further conversations about the need to incorporate data science education within their domain.

During the preparation of this article, we collected qualitative feedback from Collaboratory Fellows in the form of six interviews and 23 surveys. We also conducted two Collaboratory Fellows gatherings during which we led structured discussions on best practices in data science education, the Collaboratory's impacts, and the future of the program. Furthermore, we analyzed the contents of the 21 winning Collaboratory proposals, which detailed the need for the proposed course or courses, the aims of the project, its target student population, evaluation methods, the composition of its instructional partnership or team, and a 3-year budget. We also collected student evaluations from 11 Collaboratory courses and enrollment data for all of the courses taught so far. Readers may refer to Appendix Table A1 for data on individual Collaboratory projects. In this article, we share insights and lessons learned by the Collaboratory community collectively.

Collaboratory Fellows have been able to push forward their plans and overcome financial and administrative barriers to transdisciplinary course development. In addition to funds that specifically support curriculum development, the Collaboratory facilitates this process by providing the Fellows with institutional backing, an experience-sharing community, and the advantage of a 'stamp of approval' from Columbia Entrepreneurship and the Data Science Institute. Through the Collaboratory, Fellows are empowered to dream, imagine, prototype, and test the best pedagogical model for the data science educational needs of their students.

The Collaboratory can be thought of as an aspen grove. Aspens are noted for their ability to reproduce through root sprouting, with the potential for an entire aspen grove to grow through shared root structures (*How Aspens Grow: Quaking Aspen (Populus tremuloides)*, n.d.). The Collaboratory stems from a unifying idea and funding source, and shares the *Collaboratory Fellows* cohort structure of mutual support and collective resources (refer to Figure 1). It consistently drives the creation of new courses that serve students throughout the many domains of the university, just like an aspen grove sending out new shoots that grow into trees, quickly covering large landmasses with a beautiful canopy.

Figure 1. Collaboratory at Columbia: An aspen grove of data science education.

2. Data Science Education: Assumptions and Goals

Articulated or not, an institution's leadership faces a number of considerations when deciding whether to revamp curricula or adopt a particular model in order to expand data science education. First, it is useful to consider the guiding assumptions and goals behind a push for expanded data science education, and how a given curricular model might answer these.

First is an assumption that developing data science education requires proactive efforts (Anderson et al., 2014; Cleveland, 2001; Lazowska, 2018; Moore-Sloan Data Science Environments, 2018; NASEM, 2018; Van Dusen et al., 2019; Wing et al., 2018). On a macro level, difficulties include the cost and expertise needed to develop data science offerings. At the level of faculty and student experience, there is a need to shift mindsets around data science and make it a field that is accessible and inclusive of students from different demographic and academic backgrounds (NASEM, 2018).

Second is an assumption that data science education should be collaborative and transdisciplinary (Anderson et al., 2014; Blei & Smyth, 2017; Cleveland, 2001; Moore-Sloan Data Science Environments, 2018; NASEM, 2018; Pawlicka, 2017; Van Dusen et al., 2019). Data science analyzes data in a particular domain, and should not be divorced from the insights of the fields that traditionally study and practice within that domain (Anderson et al., 2014; Lewis-Strickland, 2018; Pawlicka, 2017). However, transdisciplinary collaboration

must bridge structural and bureaucratic divides (Lazowska, 2018), not to mention the seemingly incompatible ‘cultures’ of particular disciplines (Anderson et al., 2014; Irwin et al., 2018). What does truly collaborative and transdisciplinary data science look like? Even within transdisciplinary collaborations, there are many potential constellations of power dynamics, such as the flow and direction of information and expertise, as well as the clout to enact change in the world (Pawlicka, 2017). Furthermore, advice and support for students from different academic backgrounds can be challenging in transdisciplinary classrooms (Anderson et al., 2014).

Third is a growing push within universities and the wider society for data science to be taught and conducted with attention to ethical implications (Anderson et al., 2014; NASEM, 2018; Van Dusen et al., 2019; Wing, 2018; Wing et al., 2018; Zook et al., 2017). Even institutions at the forefront of data science are wrestling with how to meet this goal. How should ethics be made integral to data science coursework? Is ‘exposure to ethics’ enough, or should students be challenged to design technical solutions to ethical problems? In other words, how might we push students from a position of whistleblower to problem-solver in a technical sense? As institutions consider implementing a particular model of data science education, their approach to ethics may be at the forefront of this evaluation.

Fourth, data science education must meet the needs of institutions working in different fields and serving different student populations, and thus might require a variety of approaches (Moore-Sloan Data Science Environments, 2018; NASEM, 2018; Wing et al., 2018; Wing & Banks, 2019). One priority may be meeting industry demand for data science expertise, including the local and regional businesses that enrich a region. For any given student population, the question ‘What is an adequate or desired level of data science competency?’ must be asked—what the Moore-Sloan Data Science Environments call “depth of coverage” (Moore-Sloan Data Science Environments, 2018). In Table 1 we introduce a framework in which data science competency is broken down into four distinct levels. Within the Collaboratory, this framework has been helpful to differentiate among approaches to meet the data science needs of students across diverse fields and stages of education.

Table 1. Data science competency framework.

DATA SCIENCE COMPETENCY LEVELS	
Data literacy	☐ Understand on a theoretical level what data science is, how it is used, and how it can be relevant to a field of inquiry ☐ Understand sources of biases and ethical implications of data-driven solutions ☐ Critique the growing field of data science from the position of an informed humanities scholar, artist, social scientist, lawyer, journalist, and so on. ☐ Communicate and collaborate effectively with data scientists on shared projects
Data science skills	☐ Attain skills to work with data through coding, data visualization, etc. ☐ Master data science skills in a particular domain relevant to a student's field of work
Data scientist	☐ Build metacognitive ability to apply a broad data science skillset to solve problems, that is, data acumen (Bargagliotti et al., 2020; NASEM, 2018).
Data science researcher	☐ Achieve highly skilled capacity to carry out data science research independently ☐ Pursue a graduate or professional career in data science ☐ Make advances in cutting-edge data science technologies and techniques

3. Collaboratory at Columbia Model

The Collaboratory model offers an innovative approach to addressing the considerations described here. The model takes a proactive approach to cultivating collaborative data science education by providing seed funding to new crowd-sourced transdisciplinary educational offerings. It embeds ethics by situating data science in context. This approach also responds flexibly to disciplinary and institutional heterogeneity as the Collaboratory Fellows are the experts on the needs in their field. This flexibility extends to the needs of students from diverse domains, as the model supports coursework that spans the range, from providing basic data literacy to supporting emerging data researchers. The program scaffolds successful course development through a well-defined project life cycle and a community for the trainers.

Funding for the Collaboratory was raised through the generosity of Columbia alumni donors. Once proposals are approved by the Collaboratory academic leadership, grants are distributed through the provost's office. Grants have been made on three tiers: pilot or 'milestone' awards given to promising projects that may need further development, full awards, and supplemental funding to transition a completed in-person course to an online or hybrid format. The [2021 request for proposal \(RFP\)](#) capped full awards at a total of \$100,000, with pilot awards typically funded at half that amount. Supplemental awards after 3 years have been in the range of \$5,000 to \$25,000. A large portion of Collaboratory funds has been applied to support faculty efforts on curriculum development and piloting new courses.

3.1. Data Science Education in Context

The Collaboratory drives curricular development in context, at an appropriate data science competency level for the target audience. Rather than creating a single point of entry for data science education, the Collaboratory fosters a diverse set of new pedagogical projects incorporating data science into the study of different fields across the university. This offers a unique *meta-model* for data science education in comparison to a general purpose (i.e., one-size fits all) or foundational data science course. As such, Collaboratory curricular content range from basic data literacy to skills that can be immediately applied to students' ongoing data science research in their fields. This diversity of content also brings insights and concerns from other disciplines into the study of data science.

3.2. Crowd-Sourced Curriculum Development

The Collaboratory capitalizes on the interests and passions of faculty. Instead of dictating new courses through a top-down process, the Collaboratory uses a crowd-sourcing approach. This bottom-up structure has allowed a broad range of educational approaches and experimentation to take place. This is fitting, as the pedagogy developed meets the needs of heterogeneous learners, including undergraduate and graduate students across diverse academic programs, and even other faculty, as well as medical practitioners, and so on, who have taken advantage of these offerings. In academic contexts, where commitment to innovative teaching is not always recognized as highly as innovative research, the Collaboratory upended this dynamic by incentivizing and bestowing prestige upon cutting-edge curricular design. Examples include a flipped classroom approach (Tucker 2012), a Common Task Framework (Dohono 2017), and Peer Learning (Boud et al., 1999). A number of courses have used online collaboration platforms such as Slack and GitHub to facilitate project-based learning collaborations between students with different disciplinary expertise. Furthermore, the Collaboratory has demonstrated the value of fostering preexisting talents and interests, rather than looking toward external hiring as the solution to expanding data science coursework.

3.3. Collaboratory Project Life Cycle

By providing an injection of seed funding, the Collaboratory helps a diverse group of pedagogical projects quickly move through planning stages to student impact along a common *life cycle*. A request for proposals for

the Collaboratory Fellows Fund is posted annually, setting into motion the five stages of Collaboratory projects in Table 2.

Table 2. Five stages of collaboratory projects.

<p>RELATIONSHIP</p>	<p>First, a relationship is built between faculty interested in collaborating on a project, often across disciplines and between domain scientists and data scientists. While many of the pioneering groups were made up of faculty with <i>preexisting</i> research relationships, other relationships were developed through networking inspired by the Collaboratory call for proposals itself. If a faculty member has a project idea but does not know a colleague with complementary computational expertise, Collaboratory leadership will suggest a data scientist they can approach.</p>
<p>PROPOSAL</p>	<p>Second, a team develops a Collaboratory proposal that sets out a pedagogical project. This proposal outlines the need for the course, target student populations, aims and outcomes, assessment strategies, the team members' credentials and contributions, and a budget within the scope of the Collaboratory's 3-year funding structure. The proposal must also envision how the developed curricula will be integrated into an academic program over the long-term. The feasibility of this plan must be supported with letters from department chairs and deans who have committed to continue offering the course after Collaboratory funding ends.</p>
<p>PLANNING AND SCOPING</p>	<p>In the third stage, upon receiving a Collaboratory Fellowship, each team goes through a planning and scoping period. They plan their course(s) learning outcomes, targeting a particular data science competency level based on the student population(s) served. They develop a syllabus and accompanying tools. This process benefits from expertise shared among the community of Collaboratory Fellows. In the case of traditional coursework, the syllabus must be approved through the normal channels and added to the course registry. The third stage of planning typically takes 6 months to 1 year.</p>

<p>IMPLEMENTATION AND PILOTING</p>	<p>In the fourth stage, the course is piloted. Collaboratory Fellows are expected to pilot their course by year 2 of the funding period. Many student cohorts include a mix of students with computational backgrounds and those with domain backgrounds in the social sciences, humanities, arts, hard sciences, and so on. As such, how to support and bridge these domains is a learning experience during the pilot round of the course.</p>
<p>DEPLOYMENT AND SUSTAINING</p>	<p>Fifth comes the evaluation of the pilot round and reimplementing, incorporating lessons learned. This is a cyclical process in which each iteration of the course informs modifications made in the next round.</p>

3.4. Proposal Development and Evaluation

The Collaboratory [RFP](#) requires a minimal set of application materials to encourage creative proposal design while avoiding too much of a burden for interested faculty. Collaboratory proposals are evaluated based on the following rubrics that address their data science education value creation, feasibility and deliverables, evaluation and sustainability, and their approach to diversity, equity, and inclusion:

- Proposal identifies need for data science/computational training in the said discipline that is not currently met by any existing curricular offerings;
- Proposal identifies pedagogical, instructional, and delivery methods that will be effective for the learners targeted in the proposal;
- Proposal identifies a pair (or more) of instructors that include one from the following: domain expertise and data science/computational expertise;
- Proposal outlines a realistically achievable outcome that also includes a letter of commitment by the dean(s) of the school(s) for which the training will be housed;
- Proposal outlines an appropriate evaluation plan;
- Proposal demonstrates the way(s) in which the new offering has potential to contribute to the goal of embedding data literacy across disciplines;
- Proposal demonstrates an evidence-based approach to serving the Collaboratory's values on diversity, equity, and inclusion, both in terms of the topics that will be taught, as well as instructional approaches.

3.5. A Community for the Trainers

Across great disciplinary diversity, the Collaboratory has created a community working with a common sense of purpose. The Collaboratory’s decentralized approach to pedagogical innovation is made internally strong through its cohort model. A new cohort of Collaboratory projects is funded each year. As such, well-established projects are in a position to offer resources and insights to projects in earlier stages. At regular gatherings, Fellows share best practices and resources developed in their classrooms. For example, the team behind “What is a Book for the 21st Century?” (Appendix Table A1) worked with the Center for Teaching and Learning to develop a [Digital Literacy Competency Calculator](#) (n.d.) that has since informed other Collaboratory Fellow’s assessment strategies. Fellows also discuss challenges they face and help identify solutions. Confronting common challenges such as how to register transdisciplinary courses, the Collaboratory community has strength in numbers to advocate for flexible institutional structures that support transdisciplinary collaboration.

The Collaboratory community also celebrates Fellows’ and students’ achievements. These achievements are often directed toward wider audiences, aimed at sharing new understandings and tools. For example, the project “What is a Book for the 21st Century?” produced a [minimal digital edition of a 16th-century French manuscript](#) documenting an artisan’s experimental techniques (“Engineering the Future of Cultural Preservation,” 2019; “*Making and Knowing Project Launches Digital Edition of 16th Century Book of Secrets*,” 2020). The team behind “Data: Past, Present, Future” (Appendix Table A4) made their [syllabus and course materials open source](#), and Chris Wiggins presented a series of lectures on their teaching approach (Florida, 2019). Students in “Multilingual Technologies and Language Diversity” (Appendix Table A1) shared their essays publicly through the Columbia’s Institute for Comparative Literature and Society’s “[Explorations in Global Language Justice](#)” blog. Two students in “The Search for Meaning in Big Data” (Appendix Table A2) won first place in the American Geophysical Union’s Data Visualization and Storytelling contest for their [film project](#), and they are among a number of students whose films are used to teach seismology at other institutions.

Collaboratory leadership works hand-in-hand with the broader Collaboratory community to realize its [aims and impacts](#). For example, current Fellows with field-specific expertise evaluate each new round of proposals, offering detailed feedback to teams based on their own experience and insights. This serves to guide teams who will be funded toward a successful outcome, as well as to guide teams that were not selected toward ways they can strengthen their proposal for future consideration. Furthermore, Collaboratory leadership frequently connects projects with skilled Teaching Assistants from the Data Science Institute. The Collaboratory also helps with visibility so that students and other faculty are aware of these unique course offerings. Fellows have additionally contributed vital insights toward mapping out areas in the curriculum that are ripe for a Collaboratory course, as well as setting the agenda for the Collaboratory’s thematic focus in future years. Be it in social gatherings, evaluating proposals, or assessment activities such as the ones that fueled this article,

whenever the Collaboratory community comes together, new relationships are seeded, resulting in new collaborations in education leadership and research.

4. Value Creation Through the Collaboratory

What value has the Collaboratory created for Columbia and the wider society? We know from faculty feedback and student evaluations that Collaboratory learning has led to exciting new courses, enhanced capstone projects, new research strands, journalistic techniques applied by alumni to hold power to account, and more. In a number of cases, such as the courses developed in Columbia’s Business and Journalism schools, Collaboratory courses are recognized as cutting-edge nationally. In the Collaboratory’s first 5 years, we have reached 4,170 students through 21 funded projects. Even more exciting is that the number of students Collaboratory courses reach annually will continue to grow as recently funded projects are implemented. The program’s reach is an incredible value for the funds invested, particularly as the Collaboratory capitalizes on preexisting strengths and transdisciplinary potential within our university’s context. We hope that other institutions will find facets of our model useful in their own efforts to embed data science across the curriculum.

4.1. Categories of Curricular Models

A commonly echoed sentiment among Collaboratory Fellows was appreciation for funds not only to launch innovative curricula, but also to *design* it. Many Fellows allocated a portion of their Collaboratory budget toward buyout of other teaching duties or summer salary, giving them time for a deep dive into curricular design. A number of projects consulted with Columbia’s Center for Teaching and Learning on instruction best practices. Others surveyed professionals in the field, or gathered insights from the Collaboratory community. Out of this rich process, a number of patterns in curricular approach have emerged. While each Collaboratory project is unique and there is not space to do justice to each one, the following course categories may serve as guideposts for other institutions.

- First is a course in which a data science approach to a particular discipline is addressed. Collaboratory courses bringing introductory, ‘toe in the water’ data science literacy and/or skills to social work, nursing, dentistry, real estate, and business have all been developed. In many cases, the Collaboratory has served a capacity-building purpose in which a course that might never otherwise have been developed is given a foundation of funding and data science expertise from outside a department, college, or school.
- Second is a two-way model in which data science informs a traditional domain and that domain also informs data science. These courses are built upon transdisciplinary collaboration between both the instructional team and among a mixed cohort of students. This approach fosters cross-pollination and thrives on project-based collaboration. Many Collaboratory offerings in this category are designed to train the next generation of a transdisciplinary research workforce. These courses demonstrate and teach transdisciplinary research skills in a very practical and specific way, rather than a theoretical way. An example of this is “Multilingual Technologies and Language Diversity” (Appendix Table A1). This course brings together humanities and

computer science students to take a critical, historical look at why minoritized and Indigenous languages are digitally disadvantaged, then build the Natural Language Processing skills to redress these gaps through project-based learning.

- A third model is an add-on or bootcamp/workshop approach in which students, practitioners, and/or faculty are provided with a needed skill set in a short but intensive voluntary session. For example, the “Points Unknown” project offered journalism students nights-and-weekends workshops on data mapping skills (Appendix Table A6). Alongside its formal course, the “In Vivo Magnetic Resonance Spectroscopy” project offered a short workshop on this topic to medical residents and faculty (Appendix Table A5).
- A fourth model is a department or school that uses the Collaboratory as a ‘kick’ to rapidly develop an array of coursework. For example, Columbia’s Business School launched a series of data-intensive coursework within 12 months of gaining Collaboratory support, which would otherwise have taken 4 years to roll out. Similarly, the School of Public Health is using Collaboratory support to jumpstart a new master's-level program in Environmental Health Data Sciences as part of the “Integrating Data Science into Environmental Health Sciences’ Curricula” project (Appendix Table A5). Other examples such as “Next Gen Cognitive Neuroscientists,” which developed a suite of neuroscience courses approaching the field from computational and statistical perspectives, and the recently-funded project “Accessible and Inclusive Data Capture and Display,” which is developing a new course and revamping three existing courses, also fall into this category.

4.2. Flexible Co-teaching Arrangements

Collaboratory teams take a variety of approaches to collaborative curriculum design and co-teaching. In some cases, Fellows split teaching responsibilities 50/50, while in other cases one Fellow takes the lead and is supported by guest lectures by one or more additional Fellows. Some proposals request funding to create video lectures by one of the instructors, so that the course can be primarily managed by a solo instructor in a flipped classroom fashion. Considering all these factors, it is difficult to generalize about the workload for instructors. Of course, it always takes special effort to launch a new course, which is why the Collaboratory model is so helpful as it provides support not only for teaching but also course development. That said, many Fellows noted how enjoyable co-teaching is!

4.3. Virtuous Cycle: Transdisciplinary Teaching-Research Collaboration

Collaboratory courses engage with data science in the context of other fields from the get-go, rather than as the outgrowth of a foundational data science education. They stem from partnerships that are developed between a domain expert in a field such as dentistry, social work, or the arts, and a data scientist who is called upon to put their skills in service of this domain. The domain expert typically brings the vision and drive to develop a new curriculum on a particular topic. Often, this is because they see the urgency of an issue, and are motivated to launch a course for the benefit of their students and field. The data scientist partner is often motivated by challenges in the domain field where data science approaches can lead to advancement.

By funding curricular design, the Collaboratory amplifies preexisting interest in collaborating across disciplines. A large number of Collaboratory Fellows commented that the highlight of their experience was working with colleagues and students across disciplinary divides. Some commented on the joy of seeing data science tackled from so many angles, while others appreciated the opportunity to see their expertise applied within an unfamiliar field. Many Collaboratory Fellows emphasized how much they enjoyed working with students from a range of academic disciplines.

A number of Collaboratory projects leveraged prior research collaborations between Collaboratory Fellows. Instructors' joint research enriches course material and models transdisciplinary collaboration. Feedback from Fellows demonstrates that the process of developing and teaching new coursework has deepened preexisting research relationships, and in a number of cases led to the development of new collaborative research. For example, an MFA student from the course "Interpreting Urban Environmental Data" (Appendix Table A3) drew on the course for her final project, while other students developed transdisciplinary research projects with faculty that continued beyond the course. In another example, Fellows John Paisley and Ben Holtzman first collaborated on a research project, then decided to partner on the Collaboratory project "The Search for Meaning in Big Data" (Appendix Table A2). As part of this project, they built the "Spatial Sound Lab" at Columbia's Computer Music Center, a 32-channel spatial sound system coupled to virtual reality systems, designed for data exploration and artistic purposes. The lab has enhanced its own research activities as well as its students' research opportunities. This team's multifaceted collaboration across the domains of engineering, art, earth sciences, and data science is a fitting example of the virtuous cycle the Collaboratory has nurtured between teaching and research across wide disciplinary divides.

4.4. Diverse Curriculum for Diverse Students

Collaboratory courses bring data science education to students from across the traditional domains of the university. Student feedback demonstrates that students who would otherwise be too intimidated to take a data science class have found the courage to approach the subject through the gateway of a Collaboratory course firmly rooted within a familiar field. We have seen that the resulting diversity of perspectives in Collaboratory courses leads to rich ethics discussions and new innovations in problem-solving. In turn, the Collaboratory is diversifying the areas of domain expertise entering the data science pipeline; numerous students commented in course evaluations that they had developed further plans to study data science or apply it to their research, thanks to a Collaboratory course. In terms of demographic diversity, many Collaboratory courses explicitly aim to promote greater diversity, equity, and inclusion within the field of data science, and this has recently become a requirement of all proposals.

4.5. Embedded Ethics

Our community has developed the insight that data science education firmly rooted in an external discipline allows ethical concerns to emerge intentionally and organically. This is because attention to the motivations

behind data collection and use, attempts to identify truth, and wariness about the intrusion of bias are foundational concerns of practitioners in fields such as the social sciences, humanities, journalism, and social work. A significant portion of Collaboratory projects emphasize ethics and equity, including “Data Sciences for Social Good,” “Interrogating Justice and Ethics in Digital Health,” “Multilingual Technologies and Language Diversity,” “Introduction to NYC Health Disparities using Data Science,” “Data: Past, Present, Future,” and “Points Unknown: New frameworks for Investigation and Creative Expression through Mapping” (see Appendix).

One of the stated goals of many Collaboratory courses is to achieve reverse polarity by bringing different domains into data science. That is, instead of or in addition to teaching data science skills within a traditional domain, they cast a critical eye on the practice of data science itself, embedded as it is within specific histories and power structures. For example, Fellows Chris Wiggins and Matt Jones’s breakthrough when teaching “Data: Past, Present, Future,” was realizing that their class was about “truth and power” and that arranging the course chronologically allowed the power dynamics that shape how data is collected, modeled, and utilized to emerge (Florida, 2019).

Student feedback reflected the uniquely deep engagement with ethics that Collaboratory courses provide. Students reflected upon their appreciation for courses in which ethics are an integral part of coursework, rather than a superficial add-on to core topics. Furthermore, the emphasis on data science skills within Collaboratory courses enables students to transition from a position of ethics whistleblowers to ethically attuned problem-solvers. This well-rounded skill set is highly appreciated by students who have gone on to use newfound expertise in their research, journalism, or other professional practice. For example, journalist alumni of “Points Unknown” (Appendix Table A6) are now utilizing their skills in data mapping to highlight justice and equity issues at *The Marshall Project* and *The New York Times*.

4.6. Agile Response to Industry Trends

The Collaboratory supports curricular development that contributes to both general education goals as part of Columbia’s liberal arts mission, as well as workforce development offered by Columbia’s various professional schools. Columbia is located in a metropolitan area with a great emphasis on data science innovation and industry, and the Collaboratory benefits from and feeds into that ecosystem. The opening paragraph of the RFP states, “Scores of professional and research areas are increasingly becoming data-driven enterprises. To prepare for a data-rich world, students and future leaders need to integrate data science into their traditional areas of study.” A strength of the Collaboratory model is that many proposals are being driven by successful and popular academic programs that are motivated by business and career trends. The Collaboratory does not prescribe new courses; rather the crowd-sourcing model allows faculty to agilely respond to industry needs. In fact, market trends are part of the first evaluation criterion used to assess Collaboratory proposals (see section 3.4 above).

As a result, many Collaboratory projects were designed to meet industry needs through innovative coursework. Examples include courses that teach professional competencies in digital scholarship required of humanities graduate students entering the academic job market, to capstone projects carried out in the School of International and Public Affairs with institutional partners in the business, governmental, and nongovernmental sector, to a suite of Columbia Business School courses designed to position students for professions in business-specific data analytics or to launch a new wave of data-driven companies, to modules that introduce journalists to data mapping techniques. Other examples include projects conducted in the Graduate School of Architecture, Planning, and Preservation; Mailman School of Public Health; School of Nursing; School of Social Work; and the Journalism School.

5. Facing Challenges, Moving Forward

5.1. Challenges

The success of Collaboratory projects has not come without common challenges. The primary challenge is navigating administrative systems, such as course registration, that are built around distinct departments and schools rather than transdisciplinary partnerships, co-taught courses, and shared budgets. In other cases it has been a logistical challenge to meet the needs of targeted trainees who cannot take traditional coursework, such as physicians in training at Columbia's Medical Center. Navigating these processes is often iterative and may take several years to refine the most student-friendly way to cross-register a new course across two different schools, for example. Frequently, the prestige of Collaboratory funding and backing from Columbia Entrepreneurship and the Data Science Institute has been helpful in facilitating these processes. The Collaboratory program has stimulated discussion about structural changes that can facilitate all forms of transdisciplinary partnerships in the future.

Another challenge lies in balancing the needs of mixed cohorts of students, which many Collaboratory courses do. In this context, some students need introductory computational training, while others may already have competence at the level of data science researchers. Some courses offer additional labs or bootcamps to provide basic data science literacy and skills to those who need them. Other courses employ additional Teaching Assistants to boost one-on-one support. Fellows have also identified project-based learning as a best practice when working with mixed cohorts, with collaboration taking place across domains of expertise. In some cases students contribute to their team *within* their area of expertise, while in other cases students work against the grain of their existing expertise in order to build new skill sets. In this case a partnership might be assigned in which a humanities major writes code while a computational student writes an accompanying essay providing historical and ethical context, and both partners advise one another.

Both Collaboratory Fellows and students note the challenge inherent in teaching computational skills to newcomers with time to spare in the semester to also develop projects at an interesting depth. Revisiting the

data science competency framework proposed earlier in order to develop realistic expectations for course outcomes is one approach our community has identified for managing this challenge.

Most Collaboratory courses are co-taught, and coordinating among multiple instructors can present challenges; that said, the majority of responses make it clear that the benefits of co-teaching far outweigh the challenges. Integrating very different disciplines and learning enough about a new domain has been a challenge for some Collaboratory Fellows. In a handful of cases, faculty turnover or the need to recruit specific expertise has temporarily impeded a project. Long term, recruiting faculty with competencies that match existing Collaboratory coursework has the potential to further seed data science across the curriculum.

Some Collaboratory courses have faced sustainability challenges after Collaboratory funds are exhausted. This has been due to changes in leadership or funding priorities at the level of a college or school. The Collaboratory has been able to offer an additional round of financial support in a few such cases. However, the majority of Collaboratory courses in years 3–5 have been successfully implemented multiple times. The sustainability of courses in the curriculum continues to be the Collaboratory community's number one priority.

5.2. Areas for Growth

After collecting 5 years of best practices, we have identified several directions that hold great potential for the future of the Collaboratory program and the data science education community more broadly. Several projects are in the works. First is a shared collection of assessment instruments for various topics in data science. Second is a repository of syllabi, which we plan to make open to the public upon completion via the [Data Science Institute website](#). A recognition of the importance of shared instructional materials is not new, but our plan hopes to optimize accessibility by organizing modules within syllabi along two axes: 1) pure mathematics foundations to applied, domain-specific topics and 2) modules for self-teaching versus modules designed for a standard course format. Third is a repository of common data projects with well-designed scaffolding and tutorials that can serve the needs of learners in different courses.

Additionally, we have begun to fund efforts to shift Collaboratory courses into online or hybrid formats, utilizing best practices in this domain of instruction. The COVID-19 pandemic spurred recognition of the importance of robust approaches to online teaching. Beyond the pandemic, this will remain a long-term strategy to make Collaboratory courses sustainable and accessible. Since we have developed a community of innovative data science educators able to effectively reach learners from many different disciplines, we are also considering how we can utilize digital approaches to reach learners outside the university's walls.

Thematically, we are moving toward funding projects with an explicit focus on increasing the gender, racial, and class diversity within the data science pipeline as well as projects that support antiracist practice. Starting in 2021, diversity, equity, and inclusion (DEI) have been integrated into how projects and fellows are selected. This includes a reconceptualization of who is a domain expert, broadening not just who has access to tools, but who is building them (D'Ignazio, 2020; Frey et al., 2020; Irwin et al., 2018). Projects from previous funding

rounds as well as new proposals address DEI in different ways. Proposals in response to our 2021 RFP include “Design Justice: Human-centered Design and Social Justice” and “Frontiers of Justice: Using Data to Transform Justice,” as well as a proposal from the Journalism School and the Department of Statistics that aims to develop data-driven news stories in collaboration with the diverse New York City communities that the stories represent. To further scaffold DEI efforts, we circulated a set of [resources on race and data science](#) collected by Collaboratory Fellow and Associate Director of Diversity, Equity, and Inclusion at the Data Science Institute, Desmond Patton, to faculty responding to the 2021 RFP. We are also collecting best practices in teaching for DEI such as those presented in Columbia’s “[Anti-racist Pedagogy in Practice: First Steps](#)” resources, and will organize a Fellows meeting to share and discuss. We also plan to make curriculum design for diversity a topic for a future Collaboratory Fellows meeting. We hope that these efforts will support Fellows’ success in embedding DEI into their projects for the benefit of students and society.

6. Conclusion

How does the Collaboratory relate to the founding assumptions behind data science education outlined earlier? First, it breaks the difficult task of developing an interdisciplinary slate of data science education offerings into mutually supportive, crowd-sourced projects. These projects are fueled not by a top-down mandate, but rather by collaborations among faculty that draw on their preexisting expertise and shared interests. Second, it creates spaces where students from different fields can intermingle, feel comfortable, and succeed by collaborating across diverse skill sets. Third, it flexibly facilitates a match between a particular field and the appropriate approach to data science integration, including the data science competency level needed for the target student population. Fourth, it intentionally and organically includes ethics, as the pedagogy developed is grounded in the specificity of particular domains. In fact, it pushes students to move past identifying ethical quandaries to become problem-solvers who can develop data-enabled solutions or recognize when other approaches are necessary. Fifth, through co-teaching experiences, often with mixed cohorts of students, the Collaboratory inspires transdisciplinary data science research collaborations at the level of faculty and students that extend beyond the life of a course.

It is especially important to reiterate that the Collaboratory has raised awareness of and support for data science education across the university. The Collaboratory provides seed funding, which is school/discipline agnostic, yet is dependent on the stated support of department chairs and deans to maintain the sustainability of courses over the long term. Over the past 5 years this has built a broad base of understanding among university leadership regarding the need to introduce data science to students across all domains. As a result, more data science efforts are underway across Columbia’s campuses.

Like an aspen grove, the Collaboratory has fostered a diverse and vibrant set of context-specific pedagogy, a community of instructors that nurtures transdisciplinary research, and a groundswell of student interest able to benefit from the ‘canopy’ of Collaboratory offerings (refer to Figure 1). It has also enriched the “soil” of the university, building appreciation among university leadership for the contributions data science can make

across the curriculum. We hope the flexible best practices the Collaboratory model provides will prove useful for other institutions seeking to build bottom-up, transdisciplinary data science coursework that supports their own unique context.

Disclosure Statement

Isabelle A. Zaugg, Patricia J. Culligan, Richard Witten, and Tian Zheng have no financial or non-financial disclosures to share for this article.

References

- Anderson, P., Bowring, J., McCauley, R., Pothering, G., & Starr, C. (2014). An undergraduate degree in data science: Curriculum and a decade of implementation experience. In *Proceedings of the 45th ACM Technical Symposium on Computer Science Education* (pp. 145–150). ACM. <https://doi.org/10.1145/2538862.2538936>
- Bargagliotti, A., Binder, W., Blakesley, L., Eusufzai, Z., Fitzpatrick, B., Ford, M., Huchting, K., Larson, S., Miric, N., Rovetti, R., Seal, K., & Zachariah, T. (2020). Undergraduate learning outcomes for achieving data acumen. *Journal of Statistics Education*, 28(2), 197–211. <https://doi.org/10.1080/10691898.2020.1776653>
- Baumer, B. (2015). A data science course for undergraduates: Thinking with data. *The American Statistician*, 69(4), 334–342. <https://doi.org/10.1080/00031305.2015.1081105>
- Blei, D. M., & Smyth, P. (2017). Science and data science. *Proceedings of the National Academy of Sciences*, 114(33), 8689–8692. <https://doi.org/10.1073/pnas.1702076114>
- Boud, D., Cohen, R., & Sampson, J. (1999). Peer learning and assessment. *Assessment & Evaluation in Higher Education*, 24(4), 413–426. <https://doi.org/10.1080/0260293990240405>
- Chun, W. H. K., & Rhody, L. M. (2014). Working the digital humanities: Uncovering shadows between the dark and the light. *Differences*, 25(1), 1–25. <https://doi.org/10.1215/10407391-2419985>
- Cleveland, W. S. (2001). Data science: An action plan for expanding the technical areas of the field of statistics. *International Statistical Review / Revue Internationale de Statistique*, 69(1), 21–26. JSTOR. <https://doi.org/10.2307/1403527>
- Collaboratory at Columbia University. (n.d.). Preparing Tomorrow’s Leaders for a Data Rich World. Retrieved January 21, 2021, from <http://entrepreneurship.columbia.edu/collaboratory/>
- Data Sciences Education Rapid Action Team. (2015). *Data sciences @ Berkeley: The undergraduate experience*.

D'Ignazio, C. (2020, February 21). 5 questions on data and context with Desmond Patton. *Data Feminism*. <https://medium.com/data-feminism/5-questions-on-data-and-context-with-desmond-patton-5a09661cbbc6>

Digital Literacy Competency Calculator (DLCC). (n.d.). Digital Literacy for Instructional Practices Program, Columbia University Center for Teaching and Learning and Columbia University Libraries. Retrieved February 4, 2021, from <https://ccnmtl.github.io/digital-literacy/#contributors>

Donoho, D. (2017). 50 years of data science. *Journal of Computational and Graphical Statistics*, 26(4), 745–766. <https://doi.org/10.1080/10618600.2017.1384734>

Engineering the Future of Cultural Preservation. (2019, December 13). *Columbia Engineering Magazine*. <https://magazine.engineering.columbia.edu/web-exclusive-engineering-future-cultural-preservation>

Explorations in Global Language Justice. (n.d.). Columbia University, Institute for Comparative Literature and Society. Retrieved February 4, 2021, from <https://languagejustice.wordpress.com/>

Florida, R. (2019, July 25). DSI's Chris Wiggins discusses the history and ethics of data at Princeton, Facebook and Berkeley. *The Data Science Institute at Columbia University*. <https://datascience.columbia.edu/news/2019/dsis-chris-wiggins-discusses-the-history-and-ethics-of-data-at-princeton-facebook-and-berkeley/>

Frey, W. R., Patton, D. U., Gaskell, M. B., & McGregor, K. A. (2020). Artificial intelligence and inclusion: Formerly gang-involved youth as domain experts for analyzing unstructured Twitter data. *Social Science Computer Review*, 38(1), 42–56. <https://doi.org/10.1177/0894439318788314>

Hardin, J., Hoerl, R., Horton, N. J., & Nolan, D. (2015). Data science in statistics curricula: Preparing students to “think with data.” *arXiv*. <https://doi.org/10.48550/arXiv.1410.3127>

How Aspens Grow: Quaking Aspen (Populus tremuloides). (n.d.). U.S.D.A. and U.S. Forest Service. <https://www.fs.fed.us/wildflowers/beauty/aspen/grow.shtml>

Irwin, E. G., Culligan, P. J., Fischer-Kowalski, M., Law, K. L., Murtugudde, R., & Pfirman, S. (2018). Bridging barriers to advance global sustainability. *Nature Sustainability*, 1(7), 324–326. <https://doi.org/10.1038/s41893-018-0085-1>

King, G. (2011). Ensuring the data-rich future of the social sciences. *Science*, 331(6018), 719–721. <https://doi.org/10.1126/science.1197872>

Lazowska, E. (2018, March 4). How to encourage data-driven discovery. *The Chronicle of Higher Education*. <https://www.chronicle.com/article/Advice-How-to-Encourage/242675>

Lewis-Strickland, K. (2018, May 31). Data education—Inclusivity is the word. *South Big Data Innovation Hub*. <https://southbigdatahub.org/2018/05/31/data-education-inclusivity-is-the-word/>

Lue, R. A. (2019). Data science as a foundation for inclusive learning. *Harvard Data Science Review*, 1(2). <https://doi.org/10.1162/99608f92.c9267215>

Making and Knowing Project Launches Digital Edition of 16th Century Book of Secrets. (2020, March 5). Henry Luce Foundation. <https://www.hluce.org/news/articles/making-and-knowing-project-launches-digital-edition-16th-century-book-secrets/>

Moore-Sloan Data Science Environments: New York University, UC Berkeley, and the University of Washington. (2018). *Creating institutional change in data science*. http://msdse.org/files/Creating_Institutional_Change.pdf

National Academies of Sciences, Engineering, and Medicine. (2018). *Data science for undergraduates: Opportunities and options*. The National Academies Press. <https://doi.org/10.17226/25104>

Pawlicka, U. (2017). Data, collaboration, laboratory: Bringing concepts from science into humanities practice. *English Studies*, 98(5), 526–541. <https://doi.org/10.1080/0013838X.2017.1332022>

Rawlings-Goss, R., Cassel, L. (Boots), Cragin, M., Cramer, C., Dingle, A., Friday-Stroud, S., Herron, A., Horton, N., Inniss, T., Jordan, K., Ordóñez, P., Rudis, M., Rwebangira, R., Schmitt, K., Smith, D., & Stephens, S. (2018). *Keeping data science broad: Negotiating the digital and data divide among higher education institutions*. *Mathematics and Statistics Faculty Publications*. https://scholar.valpo.edu/math_stat_fac_pubs/64

Tucker, B. (2012). Online instruction at home frees class time for learning. *Education Next*, 12(1), 82–83. <https://www.educationnext.org/the-flipped-classroom/>

Van Dusen, E., Suen, A., Liang, A., & Bhatnagar, A. (2019). Accelerating the advancement of data science education. In *Proceedings of the 18th Python in Science Conference* (Article 4). <https://doi.org/10.25080/Majora-7ddc1dd1-000>

West, S. M., Whittaker, M., & Crawford, K. (2019). *Discriminating systems: Gender, race, and power in AI*. AI Now Institute. <https://ainowinstitute.org/discriminatingystems.pdf>

Wing, J. M. (2018). Data for good: Abstract. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (p. 4). ACM. <https://doi.org/10.1145/3219819.3219942>

Wing, J. M., & Banks, D. (2019). Highlights of the Inaugural Data Science Leadership Summit. *Harvard Data Science Review*, 1(2). <https://doi.org/10.1162/99608f92.e45fcb79>

Wing, J. M., Janeja, V. P., Kloefkorn, T., & Erickson, L. C. (2018). *Data Science Leadership Summit: Summary report* [Technical Report]. National Science Foundation.

Yan, D., & Davis, G. E. (2019). A first course in data science. *Journal of Statistics Education*, 27(2), 99–109. <https://doi.org/10.1080/10691898.2019.1623136>

Zook, M., Barocas, S., boyd, d., Crawford, K., Keller, E., Gangadharan, S. P., Goodman, A., Hollander, R., Koenig, B. A., Metcalf, J., Narayanan, A., Nelson, A., & Pasquale, F. (2017). Ten simple rules for responsible big data research. *PLOS Computational Biology*, 13(3), Article e1005399. <https://doi.org/10.1371/journal.pcbi.1005399>

Appendix

Appendix A. Collaboratory Proposals, 2016–2020

Table A1. Collaboratory proposals in HUMANITIES.

Collaboratory Fellows	Schools/Depts	Results of project	Start Date	Current Stage	Students impacted through Fall 2020
What Is a Book for the 21st Century? Transforming Texts: Computational Approaches to Text Analysis & Visualization (https://entrepreneurship.columbia.edu/collaboratory/book-21st-century/?resource_ID=12306)					
Pamela Smith	Dept of History, Faculty of Arts and Sciences, Center for Science and Society	4 co-taught Humanities/Computer Science courses, publication of a critical digital edition of a 16th century manuscript , develop of the Digital Literacy Competency Calculator	2016/2017	Multiple years taught	112
Steven Feiner	Dept of Computer Science, School of Engineering and Applied Science				
Terence Catapano	Columbia Libraries				
Tianna Uchacz	Dept of History, Arts and Sciences				
Dennis Tenen	Dept of English and Comparative Literature, Division of Humanities, Arts and Sciences				
Neural Networks: Computational and Philosophical Perspectives					

John Morrison	Dept of Philosophy, Barnard College	1 course	2020	Planning	
Christos Papadimitriou	Dept of Computer Science, School of Engineering and Applied Science				
Multilingual Technologies and Language Diversity (http://entrepreneurship.columbia.edu/collaboratory/tech-and-language-diversity/?resource_ID=13834)					
Lydia H. Liu	East Asian Languages and Cultures Dept, Art & Sciences, Institute for Comparative Literature and Society	1 course, course website with open source syllabus and student projects , students' work highlighted in the Institute for Comparative Literature and Society's Explorations in Global Language Justice Blog	2018	Piloted	26
Smaranda Muresan	Computer Science, School of Engineering and Applied Science				
Isabelle Zaugg	Data Science Institute, School of Engineering and Applied Science, Institute for Comparative Literature and Society, Arts and Sciences				

Table A2. Collaboratory proposals in ARTS.

Collaboratory Fellows	Schools/Depts	Results of project	Start Date	Current Stage	Students impacted through Fall 2020
The Search for Meaning in Big Data: Patterns, Representation, and Empathy (http://entrepreneurship.columbia.edu/collaboratory/meaning-big-data-patterns-representations-empathy/?resource_ID=12315)					

Benjamin K. Holtzman	Lamont-Doherty Earth Observatory, Earth Institute, Dept of Earth and Environmental Sciences, Dept of Music, Computer Music Center, Arts and Sciences	1 course, “Data and Art” Visiting Artist Lecture Series, built Spatial Sound Lab at Computer Music Center	2017/2018	Multiple years taught	~40
Miya J. Masaoka	Visual Arts, School of the Arts				
John Paisley	Electrical Engineering, School of Engineering and Applied Science, Data Science Institute				
Arthur Paté	Lamont-Doherty Earth Observatory, Earth Institute				
Accessible and Inclusive Data Capture and Display: Creative Embedded Systems for Multi-Sensory Data Engagement					
Seth Cluett	Dept of Music, Computer Music Center, Arts and Sciences	1 new course and revamp 3 existing courses	2020	Planning	
Mark Santolucito	Dept of Computer Science, Barnard College				
Brad Garton	Dept of Music, Arts and Sciences				
Benjamin K. Holtzman	Lamont-Doherty Earth Observatory, Earth Institute, Dept of Music, Computer Music Center, Arts and Sciences				
Miya J. Masaoka	Visual Arts, School of the Arts				

Table A3. Collaboratory proposals in NATURAL SCIENCES.

Collaboratory Fellows	Schools/Depts	Results of project	Start Date	Current Stage	Students impacted through Fall 2020
Interpreting Urban Environmental Data: New York City's changing landscapes http://entrepreneurship.columbia.edu/collaboratory/programming-technology-analytics-curriculum-columbia-business-school/?resource_ID=4808					
Zoe Crossland	Anthropology, Arts and Sciences	1 course	2017	Piloted (paused due to pandemic)	8
Dorothy Peteet	Earth and Environmental Sciences, Arts and Sciences, Lamont-Doherty Earth Observatory, Earth Institute				
Nan Rothschild	Lamont-Doherty Earth Observatory, Earth Institute				
Jonathan Nichols	Lamont-Doherty Earth Observatory, Earth Institute				

Table A4. Collaboratory proposals in SOCIAL SCIENCES.

Collaboratory Fellows	Schools/Depts	Results of project	Start Date	Current Stage	Students impacted through Fall 2020
Computing in Context: Public Policy (http://entrepreneurship.columbia.edu/collaboratory/computational-literacy-public-policy-collaboration-sipa-seas/?resource_ID=4805)					

Merit E. Janow	School of International and Public Affairs	1 course and capstone projects	2016	Multiple years taught	248
Dan McIntyre	School of International and Public Affairs				
Adam Cannon	Computer Science, School of Engineering and Applied Sciences				
Gregory Falco	School of Engineering and Applied Affairs				
Data: Past, Present, Future (http://entrepreneurship.columbia.edu/collaboratory/data-past-present-future-2/?resource_ID=12278)					
Chris Wiggins	Applied Physics and Applied Math, School of Engineering and Applied Science	1 course, open source syllabus and course materials	2016/2017	Multiple years taught	204
Matthew L. Jones	History Dept, Arts and Sciences				

Table A5. Collaboratory proposals in MEDICINE AND HEALTH.

Collaboratory Fellows	Schools/Depts	Results of project	Start Date	Current Stage	Students impacted through Fall 2020
DDS Squared: Digest of Data Science (DDS) for Doctors of Dental Surgery (DDS) (http://entrepreneurship.columbia.edu/collaboratory/data-science-dental-surgery/?resource_ID=12311)					
Letty Moss-Salentijn	College of Dental Medicine	1 online course	2017	Piloted (Restructuring)	6
Joseph Finkelstein	Center for Bioinformatics and Data Analytics in Oral Health, College of Dental Medicine				
Ying Wei	Dept of Biostatistics, Mailman School of Public Health				
In Vivo Magnetic Resonance Spectroscopy: From Data to Clinical Benefit (http://entrepreneurship.columbia.edu/collaboratory/data-clinical-benefit/?resource_ID=12321)					

Christoph Juchem	Dept of Biomedical Engineering, Dept of Radiology, School of Engineering and Applied Science	1 course, 1 workshop	2017	Multiple years taught	50
Lawrence S. Kegeles	Dept of Psychiatry, Dept of Radiology, Columbia University Medical Center				
Neurogenomics (http://entrepreneurship.columbia.edu/collaboratory/neurogenomics/)					
Rene Hen	Dept of Neuroscience, Dept of Psychiatry, Vagelos College of Physicians and Surgeons	1 course	2018	Multiple years taught	33
Sergey Kalachikov	Dept of Chemical Engineering, School of Engineering and Applied Science, Center for Genome Technology and Biomolecular Engineering				
Irina Morozova	Computer Lab Instructor				
Integrating Data Science into Environmental Health Sciences' Curricula (http://entrepreneurship.columbia.edu/integrating-data-science-into-environmental-health-sciences-curricula/?resource_ID=19843)					

Andrea Baccarelli	Dept of Environmental Health Sciences, Mailman School of Public Health	1 course and 2 other initiatives: Departmental evaluation of how to implement data science throughout curriculum and a Master's-level program in Environmental Health Data Sciences	2019	Course Piloted	29
Jeff Goldsmith	Dept of Biostatistics, Mailman School of Public Health				
Nina Kulacki	Dept of Environmental Health Sciences, Mailman School of Public Health				
Tiffany Sanchez	Environmental Health Sciences, Mailman School of Public Health				
Introduction to NYC Health Disparities using Data Science (http://entrepreneurship.columbia.edu/introduction-to-nyc-health-disparities-using-data-science/?resource_ID=19850)					
Mary Beth Terry	Dept of Epidemiology, Mailman School of Public Health	1 course	2019	Piloted	40
Abigail Greenleaf	Dept of Population, Family and Reproductive Health, Mailman School of Public Health				
Samantha Garbers	Dept of Population, Family and Reproductive Health, Mailman School of Public Health				
Dana March	Dept of Epidemiology, Mailman School of Public Health				
Next Gen Cognitive Neuroscientists: Suite of Human Brain Imaging Courses (http://entrepreneurship.columbia.edu/building-next-gen-of-cognitive-neuroscientists-a-suite-of-interdisciplinary-human-brain-imaging-courses/?resource_ID=19859)					

Alfredo Spagna	Dept of Psychology, Arts and Sciences	3 courses	2019	2 Courses Piloted, 1 in Planning	34
Xiaofu He	Dept of Psychiatry, Columbia University Medical Center				
Lila Davachi	Dept of Psychology, Arts and Sciences				
Nikolaus Kriegeskorte	Dept of Psychology, Arts and Sciences, Dept of Neuroscience, Zuckerman Mind Brain Behavior Institute, Vagelos School of Physicians and Surgeons				
Paul Sajda	Data Science Institute, School of Engineering and Applied Science				
Chris Baldassano	Dept of Psychology, School of Arts and Sciences				
Agnes Chang	Dept of Computer Science, School of Engineering and Applied Science				
Data Science for Better Health Outcomes: A Nursing Perspective (http://entrepreneurship.columbia.edu/data-science-for-better-health-outcomes-a-nursing-perspective/?resource_ID=19867)					
Maxim Topaz	School of Nursing, Data Science Institute, School of Engineering and Applied Science	1 course	2019	Planning	
Kathleen Mullen	School of Nursing				
Kenrick Kato	School of Nursing				
Interrogating Justice and Ethics in Digital Health					

Noemie Elhadad	Dept of Biomedical Informatics, Vagelos College of Physicians & Surgeons	1 course	2020	Planning	
Sandra Soo-Jin Lee	Dept of Medical Humanities and Ethics, Vagelos College of Physicians and Surgeons				

Table A6. Collaboratory proposals in PROFESSIONAL PROGRAMS.

Collaboratory Fellows	Schools/Depts	Results of project	Start Date	Current Stage	Students impacted through Fall 2020
Programming, Technology and Analytics Curriculum for Columbia Business School (http://entrepreneurship.columbia.edu/collaboratory/programming-technology-analytics-curriculum-columbia-business-school/?resource_ID=4808)					
Costis Maglaras	Business School	10 courses	2016	Multiple years taught	~3095
Hardeep Johar	Dept of Industrial Engineering and Operations Research, School of Engineering and Applied Science				
Wei Ke	Business School				
Points Unknown: New frameworks for Investigation and Creative Expression through Mapping (http://entrepreneurship.columbia.edu/collaboratory/points-unknown-new-frameworks-investigation-creative-expression-mapping-collaboration-j-school-gsapp/?resource_ID=4801)					
Juan Francisco Saldarriaga	Center for Spatial Research, Graduate School of Architecture, Planning & Preservation	1 course, 1 bootcamp	2016	Multiple years taught	210
Marguerite Holloway	School of Journalism				
Michael Krisch	Brown Institute, School of Journalism				
Data Science for Social Good (http://entrepreneurship.columbia.edu/collaboratory/data-science-for-social-good/?resource_ID=13830)					

Desmond U. Patton	School of Social Work	1 course	2018	Piloted	29
Tian Zheng	Dept of Statistics, School of Arts and Sciences, Data Science Institute, School of Engineering and Applied Science				
Tara Batista	School of Social Work				
Transforming Curriculum in the Real Estate Development Program (http://entrepreneurship.columbia.edu/programming-analytics-and-technology-curriculum-for-gsapp-real-estate-development-program/?resource_ID=19861)					
Patrice Derrington	Graduate School of Architecture, Planning and Preservation, Center for Urban Real Estate	1 course	2019	Piloted	24
Hardeep Johar	Dept of Industrial Engineering and Operations Research, School of Engineering and Applied Science				
Data-Driven Decision-Making Modeling and Analytics					
Yi Zhang	Dept of Industrial Engineering and Operations Research, School of Engineering and Applied Science	1 course	2020	Planning	
Tony Dear	Dept of Computer Science, School of Engineering and Applied Science

Collaboratory HDSR - Appendix Tables.pdf

126 KB

©2021 Isabelle A. Zaugg, Patricia J. Culligan, Richard Witten, and Tian Zheng. This article is licensed under a Creative Commons Attribution (CC BY 4.0) [International license](https://creativecommons.org/licenses/by/4.0/), except where otherwise indicated with respect to particular material included in the article.