SE&IS White Paper on the UCLA DataX Initiative

July 10, 2020

Summary: The School of Education and Information Studies, through its professional and academic engagements, is dedicated to the construction of civil society, investigating the foundational problems of humans living together, with the aim of addressing environmental and social injustices, including the elimination of systematic oppression. We conceive of data as a form of knowledge and continuous with other forms such as books, historical records, media, and material artifacts. We take seriously the conviction that technical issues cannot be separated from cultural ones, and that social concerns weave through every aspect of the production and use of data in current global systems; data cannot be conceived, even primarily, as a problem for technical innovation alone. We propose a series of research programs dedicated to the idea of data as a social/cultural construct, from narrow conceptions such as valid operationalization and measurement of social phenomena, to larger critiques, such as the role of data in the commercialization of public space, and to forms of instrumentalist reasoning associated with the development of modernity. We also propose two new educational programs: an undergraduate major that connects technical disciplines with a liberal arts education in the humanities and social sciences around concepts of data, and a two-year professional master's program that emphasizes the technical and ethical knowledge, skills and values of creating, preserving, accessing and use of data.

A vision for the data-enabled society is first a vision of society... The School of Education and Information Studies is driven to understand the processes of forming intellectual and emotional dispositions toward participation in genuine democracy and the improvement of community life. We are concerned with the development of the individual in social and historical contexts, the "elementary problems of human's living together" (Arendt) and the production of cultural and political forms that are just and inclusive. Our work, in social systems (such as in families, communities and occupational settings) and in formal institutions (including schools, libraries, archives and the Internet) and in the development of professionals who work in those settings (teachers, librarians, designers and scholars), is aimed at the creation of civil society. We promote community engagement through a combination of research and activism, conducting activities that are premised on partnership and participation in determining the production and use of intellectual property and responsible production of data, guided by a notion of stewardship that guards against the extraction and exploitation of data.

We take seriously the conviction that technological issues cannot be separated from cultural ones, and that social concerns weave through every aspect of the production and use of information and data in current global systems. We strive to combine technical expertise with attention to critical, self-reflexive understanding of the social costs and potential benefits across diverse constituencies. Through our research and pedagogy, we aim to make evident the ways social injustice and inequity built into data practices can be redressed in order to promote positive transformation, for example, the system of social biases built into Google, or the objectification of racist presumptions in 19th century craniometric data. Full sustainability within a globally networked environment depends on increasing social justice through greater transparency, fairness, and equitable practices. Our work focuses on power structures, particularly race, ethnicity, gender and sexuality, and forms of systematic oppression. Regimes of data and their use, we feel, have more often contributed to the reinforcement of current systems of consumption and oppression, though we are interested in their liberatory potential as well.

... and to data literacies. Literacy is the capacity of individuals and communities to not only understand the knowledge of one's culture but also develop the capacity for invention, participation and social transformation. Literacy is best understood as a pluralized concept as well, of literacies including data literacy.

Expanding beyond its traditional associations with text, literacy concepts embrace media and other cultural forms more broadly. Information literacy is a concept increasingly found in K-16 curricula (such as the Common Core) that emphasizes the ability to seek, use, and evaluate information from a variety of sources, online and traditional. A critical form of information literacy asks "why does information look like this? Why do my search results look like this? What is missing? Which voices appear in print? Who is excluded? How do I learn about the knowledge of others?" Data literacy extends these ideas as well as the concept of numeracy into the realm of data, to include not only a functional ability to create and use data, but also the ability to imagine new solutions to new problems and simultaneously understand the limitations of data and the ethical qualities of its use. A data literacy approach is expansive to include cultural and historical as well as governmental, scientific, social and commercial, and incorporates a critical approach to data that attends to the values and biases on which data are produced.

The continuity of data with other forms of knowledge. We view data as part of a larger information space that focuses on documents, records, metadata, media, material objects and other artifacts of culture. Some of the first "big" datasets were pioneered in Information Studies, including catalogs of books and journal articles, and of bibliometric data. As we built systems to acquire, preserve and provide access to information, including datasets, our efforts led to the development of new kinds of service, including libraries of digital data in a variety of formats (including maps, datasets, astronomical data, etc.), post-custodial community-based archives, research information management services, and the maintenance of digital corpora used in the humanities. We have developed a large body of research and professional practice on metadata that documents and provides access to data, metadata that itself becomes data for the assessment of collections and knowledge, and also provides the basis for mapping and linking data across cultures and communities. Importantly, we developed a body of professional standards and practices, including an ethics on privacy and intellectual freedom, the right to access and share information, and notions of epistemic and social justice, civic engagement, and participatory democracy. We have sought to understand the limits, abuses and contradictions of a data society, like the difficulty in locating public data on police officer involved homicides, the use of algorithms to control data, the racial bias in data systems otherwise characterized as neutral, and the hidden costs to labor in systems of data and media.

From data science to data studies... Data science largely assumes data as self-evident, and as self-intelligible, that is, data seems to appear out of nowhere and to enjoy independent powers of persuasion. In contrast, a data studies or data culture perspective is founded on the premise that the work of interpretation is a fundamental part of data use: at every point of the data lifecycle, from design and measurement to classification, description, standardization, statistical processing, visualization, and curation, interpretative processes must intervene in order to turn measurements into actionable data. As such, data science functions as a specific kind of "data study," one that largely negates the impact of these processes. Data studies are both about making visible the cultural work that goes into data, and exploring alternative cultural dispositions for data science itself. Given our commitments to civil society and genuine democracy, it is clear that while we use statistics in our research, ultimately we are concerned not only in the technological innovations for their own merits but as means to accomplish our studies of complex social formations and as a method of intervention to create a fairer, more just society.

... to engagements with Los Angeles. SE&IS has extensive programmatic engagements in an established base of organizations across Los Angeles. Professional schools have a technical capacity for data science, and understand the complexities of creating and using data in complex social environments, but also are uniquely positioned to assess how communities interact with data because of our installed base of community organizations distributed across many sectors in a leading global city. We also have the commitment of training professionals to work in these areas, including standardization of practice and the establishment of ethics. SE&IS is a place for the consolidation and advancement for a broadly engaging and ethical body of professional practice, something that has become an obvious deficit in the hyper-specialized and technical engineering-oriented body of practice. Our ethos is one of participation, that progress comes not in telling our partners how to do their work, but in deep engagements that respect the ethos and context of our partners, who are often themselves working as partners or leaders within broader communities.

A SE&IS plan for engaging with DataX: A data studies curriculum. The School of Education & Information Studies proudly maintains its focus on the social, cultural, policy, and ethical dimensions of

information and education. Research and educational programs on data are a logical evolution for Information Studies, and a meaningful extension for Education. Our plans for participation in DataX are ambitious: a new undergraduate major and a new two-years professional master's degree, and new activities to be integrated into a rejuvenated summer instructional programs. Our plans build on our demonstrated ability to develop joint degree programs together (e.g., the new joint undergraduate minor in information and media literacy), and will require technological support, new faculty lines, the development of new curricula and courses, and will also serve as an invitation to new research projects. However the proposal also leverages existing courses in both departments, new courses in the new joint minor, and through adapting existing courses, particularly in the MLIS program. This proposal also expects contributions from a wider array of partners, potentially including the Luskin School of Public Affairs, the Social Science and Humanities divisions in the College, and the UCLA Library. We also note that the compressed period of time for the development of this white paper prevented us from developing the full consensus and review necessary for comprehensive curricular plans.

Undergraduate major. A BA data studies major would be a joint degree program that combines technical capacity in data analysis, modeling and computational methods of classifying and machine learning, with a liberal arts curriculum on the social processes that shape the meaning and construction of data, along with dedication to understanding the human and ethical dimensions of data work. A data science curriculum without those final elements is a curriculum without a conscience, or respect for the people that shoulder a disproportionate burden of the costs and failures of a data-driven society. The data studies curriculum is built on a foundation provided by a year-long sequence of courses with zero prerequisites intended to provide the mathematical and computational foundation for data science that incorporates significant social/cultural considerations, consistent with a holistic liberal arts degree. Our plan will also draw from various elements in place, including the adaptation of courses already offered by Information Studies, in combination with re-worked courses from our minor in information and media literacy. Details of the proposal are found in the appendix.

A new professional master's degree. Our vision for a master's degree is to create an instructional program for a new kind of data professional, one with the technical capacities required for most data scientist positions, but also one attentive to the various organizational contexts in which the generation, analysis and maintenance of data occur. These positions are traditionally associated with commercial information services in Internet-based industries, where businesses have struggled with a lack of ethical understanding and social awareness. We seek to address those shortcomings, while also preparing students for a wider venue of occupational settings, including in research, education and government. Following initiatives such as federal guidelines for the preservation and re-use of data, for example, we believe data researchers of the future will also need to take on roles of stewardship of data, including preserving, documenting and providing access to collections of data. A new master's degree will be a joint program between Education and Information Studies, and based in part with modified and extended courses in the current IS MLIS program, and will be done in partnership with other elements on campus.

Research and outreach. Increasing faculty capacity and developing new academic programs will help SE&IS contribute to DataX research goals. Additionally, because SE&IS is a leader amongst UCLA academic units in its engagement with public and non-profit organizations across Los Angeles, DataX can improve the ability of SE&IS to benefit our local community. We briefly highlight a handful of centers and projects as illustrative of work that engages with and will benefit from DataX:

- **CRESST**. A research center within SE&IS.
- **Data for Democracy**. Center X (located within SE&IS) created the Data for Democracy project as part of the UCLA Centennial Initiative. The project enables Los Angeles K-12 students to collect and analyze data on critical issues such as housing inequality and access to parks. K-12 students write policy briefs on these topics, published in English and Spanish.
- **Introduction to Data Science (IDS)**. Developed in 2013 as a partnership between the Los Angeles Unified School District, Center X, and the UCLA Department of Statistics, Introduction to Data Science (IDS) is a one-year course that teaches computational and statistical thinking for secondary students, with a particular emphasis on reaching students underrepresented in STEM. The IDS data analysis labs teach K-12 students how to manage and analyze data using *R*.
- **The Community Archives Laboratory**. Under the direction of Prof. Caswell, the Community

Archives Lab at UCLA explores the ways that independent, identity-based memory organizations document, shape, and provide access to the histories of minoritized communities, with a particular emphasis on understanding their affective, political, and artistic impact. Funded by the Mellon Foundation, UCLA's Community Archives Lab has launched a three-year (2018-2021) project to provide paid internships to UCLA Information Studies graduate students at community archives throughout Southern California.

More broadly, developing the technical data methodological coursework benefit SE&IS students and the social justice oriented research and outreach projects they contribute to. For example, several SE&IS graduate students who completed a recent two-course "fundamentals of programming using R" sequence have GSR positions analyzing data from the UCLA Community School. After completing the courses, these students fundamentally improved data quality and efficiency by writing *R* scripts that create analysis variables used across many analyses. To practice the programming skills they were learning, a continuing summer reading group has prepared data and graphs for a group of CU-Boulder alumni seeking to put pressure on the university to enroll more African American students.

Appendix

Data studies: A campus partnership between Education and Information Studies

The Departments of Information Studies and Education in SE&IS are prepared to contribute considerable expertise to pedagogy in the area of data. Our research and instructional programs draw on humanistic, social science, and critical perspectives to understand the role of education and information in society, and, with this white paper, proposes to extend those paradigms to understanding the role of data in our operations and beyond, in society generally. These are combined with hands-on implementation in human-computer interaction, visualization, statistical analysis, sustainable practices and other concrete applications informed by critical issues. Our existing courses are spread across the full spectrum from undergraduate to graduate, and many could be recombined or repurposed to serve as GE courses, an undergraduate major (a joint minor is scheduled to launch in Fall 2020), a new master's program, and summer institutes to create opportunities for working professionals as well as matriculated students, conceived to serve not only those pursuing a data studies diploma, but those across the university.

The strengths of Information Studies courses are their combination of attention to social justice and diversity within the production and the use of data as a cultural, not only technical, matter. These courses are designed as fundamental introductions to issues and practices for students, scholars, and professionals. From small data within the contexts of cultural practices in community organizations and humanities projects, to large scale data challenges faced by cultural and government institutions in managing the records of the past, present, and future, our focus is on critical approaches to ownership, privacy, authenticity, validity, and sustainability.

Education brings its expertise into the training of teachers, administrators, and researchers committed to providing the knowledge to advance civic discourse. With the conviction that fundamentals of computational and statistical thinking are essential tools for all citizens, our programs have provided the basis for students and educators to advance their understanding of data production and analysis. One goal has been to prepare underrepresented populations who are entering STEM fields, but these fundamentals are relevant across disciplines. A commitment to educating students directly combines with the charge to educate educators effectively.

Campus Partners

- Luskin
- Library, esp. Data Science Center
- DH

Undergraduate Program

Our current vision is to develop elements of existing courses into year-long sequences. A list of existing courses and short descriptions is below with an outline for consolidating these into a high-impact program of offerings across the curriculum.

General Education courses

- **Data Science for Everyone**: A no prerequisites course aimed at making basic data science concepts available to the general student population beyond data studies majors.
- **Data Studies Ethics and Policy**: Fulfills Society and Culture GE requirement and undergraduate writing requirement. Aimed at data studies majors, but anyone can take this course.
- **Data Visualization and Communications**: Fulfills the Arts/Humanities GE requirement.

GE courses currently taught within Information Studies

- **IS 10: Information and Power**
- **IS 11: Introduction to Digital Methods**: A basic introduction to fundamentals of digital scholarship – data modeling to web presentation
- **IS 20: Digital Cultures**
- **IS 30: Internet and Society**

Elements from the forthcoming Undergraduate minor in Media and Information Literacy

- Introduction to Media Literacy
- Introduction to Information Literacy
- Data Ethics
- History and Practice of Information Visualization

With respect to technical training, we are in the midst of developing a year-long fundamentals of programming and data science sequence using R, targeting students with limited programming experience, to be offered at the undergraduate and graduate level. The underlying philosophy is that all students can develop strong programming through courses that provide a deep introduction to a small number of important topics rather than exposure to many topics. This sequence will be integrated into existing and proposed SE&IS programs and we anticipate it will become an essential sequence for UCLA undergraduate and graduate students who feel intimidated by courses in departments like statistics and computer science.

New courses will also draw from Education's new undergraduate major in Education and Social Transformation, which "prepares students to analyze current issues in education through a social justice lens and to emerge as effective advocates for positive change." The undergraduate major focuses on three areas of competency: histories and philosophies of education; contexts of teaching, learning, and development; and inquiry and design for learning.

Background: Proposed Minor in Data Ethics

The Department of Information Studies has also been in discussions regarding an undergraduate minor in data ethics. Major portions of these elements would likely find their way into a data studies major, or could be a minor in a data science degree program. Such a sequence of courses begins with the observation that programs and organizations that utilize big data and their algorithms are increasingly pervasive and of high consequence in society. For example, information on income, savings, and buying patterns determine whether a person is eligible for a mortgage. Where one grew up and went to school determines your eligibility for parole. Interactions in social media determine what news you see. Big data has the ability to reinforce social biases, shape and limit our civic participation, and undermine our democratic values and institutions.

An undergraduate minor on data ethics seeks to address the risks of big data and their algorithms by developing concepts, values, and professional practices dedicated to enhancing respect for human rights, open and pluralistic societies, and environmentally sustainable practice. Some of the key concepts include:

- Profiling or the re-identification of individuals
- Trust, consent, and user privacy
- Ownership of data
- Openness, transparency, and accountability
- Ethical design and auditing of data and algorithms
- Responsible innovation
- Secondary uses of data and technology
- The unforeseen and unintended consequences of big data systems
- Understanding computational systems as the locus of political, social, and cultural interaction where even truth itself is shaped and determined

Potential courses:

1. Introduction to ethics for data science
2. Data and the classification of people: data-based social formations
3. Surveillance, privacy, and data science in open societies
4. Bias in data and algorithms
5. Observation, measurement, modernity: the economization of life (historical approach)
6. Political consequences of the data society
7. Institutions and ethical data management (libraries, museums, labs, governments, universities, etc.)
8. Framing critical inquiry: data and networked science (climate, biodiversity, astronomical, data use, food studies, resilience studies, etc.)

Additional topics:

- Openness
- Liberatory data approaches
- Data, policy, and law
- Digital humanities and digital social sciences

Graduate Program

Current courses

- **IS 275: Cultural Information and Multimedia**: Works with cultural datasets, visualization, text-mining and other heuristic algorithms, cultural metaphors, interfaces, etc.
- **IS 289: AI and Society**: Explores at data issues on economic, political, and design-oriented levels.
- **IS 291C: Global Media and Information**: Explores data's economic value, platform collectives, data's impact on global movements/social movements.
- **IS 260 Description and Access**: Classification techniques and the development of cultural ontologies.
- **IS 272 HCI Human-Computer Interaction**: Basic user interface design.

Elements of professional coursework

- Knowledge extraction processes
- Trust, reliability, authenticity and issues data curation and use
- Recordkeeping metadata creation and management
- Computational approaches to archival science
- Digital preservation/digital curation
- Sustainability in information professions
- Research initiatives in audiovisual data management and preservation
- Theories on the conceptual aspects of data, systems and processes
- Evaluation of data systems and services in diverse disciplinary and community contexts

Potential summer institutes

- Digital project management
- Data ethics for professionals
- Data design and modelling for scholarly research projects
- Digital curation
- Emulation and digital workstations

Community Partnerships

Garnering external funding that yields novel insights about inequality in access to education increasingly depends on the capacity to use data science methodologies to collect new data and to analyze data in new ways.

Data for Democracy in LA

The UCLA Data for Democracy in LA project enlists our university's leading research centers as partners with K-12 teachers to enrich learning and strengthen civic discourse. Together we provide teachers and students across the city the opportunity to engage with data from UCLA research centers highlighting issues of inequality and opportunity in Los Angeles. To date we have partnered with UCLA research centers and researchers to create research briefs on access to parks, immigration and housing.

UCLA Data for Democracy seeks to enhance civic reasoning about data among K-12 students in greater Los Angeles and aspires to support mathematics learning and quantitative reasoning in LA schools. We encourage rich classroom conversations about questions such as: Where does data come from? What counts as evidence? How can I use numbers, charts, and figures to communicate what I know to others? We provide teachers with ideas for helping students grapple with data, conduct related investigations, and participate in evidence-based discussions about how to expand opportunity and deepen democracy in Los Angeles.

Introduction to Data Science

Introduction to Data Science (IDS), developed in 2013 at UCLA in partnership with the Los Angeles Unified School District, is a one-year course that teaches computational and statistical thinking for secondary students with a particular emphasis on reaching students underrepresented in STEM. The primary goal of IDS is to activate students to become "citizen data scientists", eager to address burning issues by collecting, considering, and analyzing data. Central to this goal are the data analysis labs, which teach students how to analyze and manage data using R, via the RStudio interface. The labs are supported by two additional components. The first is a series of active lessons employing an equity-driven pedagogy designed to teach fundamental concepts of statistics and computer science. The second is a set of Participatory Sensing campaigns, in which students use their mobile devices to collect data about their surroundings. Along the way, students learn to be critical and constructive users of data, and to understand the role it plays in their daily lives. In California, IDS is an A-G, program-approved status course designated as Statistics that meets a "C" mathematics requirement for high school students.

The IDS Project is the leading national provider of high school data science education materials, professional development, and technological support. By 2025, the IDS Project intends to be a center for research and development of data education tools and an advocate for educational policy change.

Data science offers unique opportunities to students; beyond providing a pathway to college, data science arms students with employable skills and tools for 21st-century-citizenship. Because technology changes rapidly and understanding data science is in demand, we need to teach sooner and reach more students, rather than wait for a privileged few to learn data science in college.

IDS was developed specifically to increase pathways to college for students belonging to groups that are underrepresented in STEM. Its aim is to prepare students to work with data in the 21st century. Students work with real data, using statistical, computational, and graphical tools for reasoning about the world.

Successful implementation of the IDS course requires significant professional development as well as a sophisticated technology infrastructure (relative to most districts' existing infrastructure). The IDS team, via UCLA's Center X, is currently supporting 15 Southern California school districts and approximately 3,500 students and teachers. The districts sign two-year partnerships with Center X. The Center provides professional development, hosts the software on its server, and provides technical support.

By the end of this year, over 12,500 students will have completed IDS since its genesis. Next year, IDS will double the number of students enrolled in the course and using its curriculum in the 2019-2020 school year. For the first time, IDS will expand beyond California to include districts in Oregon, Idaho, and New Jersey.

CRESST

C2I2

Community Archives Laboratory

EMRA Research Group

The Enrollment Management, Recruiting, & Access (EMRA) group is a UCLA-based research team founded by Ozan Jaquette, associate professor of higher education at UCLA, and Karina Salazar, assistant professor of higher education at the University of Arizona. The long-term goal of this group is to investigate the enrollment management practices of colleges and universities and examine the effect of these practices on access for underserved students.

EMRA's first project investigates the off-campus recruiting patterns of selective public and private colleges and universities, and examines whether there are any socioeconomic or racial biases in the locations that the institution chooses to visit. Various data science methods were utilized to collect, analyze, and visualize the data. For example, Python scripts were written to web-scrape admissions webpages that post off-campus recruiting events and store the raw data in a relational database. Python and R were also used to clean the data, merge them to secondary data sources on high school and community characteristics, and create analysis and visualizations. Results from this study can be found in this New York Times Op-Ed, policy brief funded by the Joyce Foundation, and interactive map.

The EMRA group is currently expanding its research to study student list purchases (e.g., universities purchasing student names and contact info from College Board and ACT) and has partnered with the Lawyers' Committee for Civil Rights Under Law to issue Freedom of Information Act (FOIA) requests to public universities for data.