SE&IS White Paper on the UCLA DataX Initiative

Gregory Leazer & Ozan Jaquette

July 7, 2020

Introduction

Here is an example of linking to Appendix A.

Some helpful resources:

- https://m-clark.github.io/Introduction-to-Rmarkdown/appendix.html
- https://rmd4sci.njtierney.com/start.html#referencing-a-section
- https://pandoc.org/MANUAL.html#heading-identifiers

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Dignissim enim sit amet venenatis urna. Massa sapien faucibus et molestie. Tincidunt eget nullam non nisi est. Enim nec dui nunc mattis enim ut. Vel elit scelerisque mauris pellentesque pulvinar pellentesque. In fermentum posuere urna nec tincidunt praesent semper. Natoque penatibus et magnis dis parturient montes. Egestas pretium aenean pharetra magna ac placerat. Malesuada proin libero nunc consequat interdum varius sit. Id interdum velit laoreet id donec ultrices tincidunt. Ut faucibus pulvinar elementum integer enim neque. Elit duis tristique sollicitudin nibh sit amet commodo nulla facilisi. Lorem mollis aliquam ut porttitor leo. Arcu non odio euismod lacinia. Nunc mi ipsum faucibus vitae aliquet. Semper eget duis at tellus at. Aliquet risus feugiat in ante metus dictum at tempor. Vulputate odio ut enim blandit.

Appendices

Appendix A: Curriculum

The Departments of Information Studies and Education in SEIS are prepared to contribute considerable expertise to pedagogy in the area of data. Our courses draw on humanistic, social science, and critical perspectives to understand the role of data in society. 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), summer institutes to create opportunities for working professionals as well as matriculated students, and courses at the graduate level that serve a broad population at UCLA.

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.

We would like to see the existing courses reach a wider population of students across the University and organize these into year-long sequences to serve a range of purposes. 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 (* to be developed or modified from an existing course):

Partners [WHERE/HOW SHOULD CURRICULUM PARTNERS BE LISTED?]

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

A1: Undergraduate major

Data Studies

A2: Undergraduate minor and/or +1 masters in data ethics

Data Ethics

Big data and their algorithms are pervasive and of high consequence. Information on your income, savings, and buying patterns determine whether you are eligible for a mortgage. Where you grew up and went to school determines your eligibility for parole. Your 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

Minor in Information and Media Literacy

Description and emphasize new minor as a demonstration of departments working together. (Comment from Jeff: Anne will have text).

Introduction to Critical Media Literacy

Critical Digital Media Literacies

A2.1: GE courses [NOT SURE IF THIS IS MEANT TO BE SUBHEADING A2.1?]

[ARE BELOW 3 ALSO COURSES OR JUST CATEGORIES?]

Arts and Humanities

Society and Culture

Scientific Inquiry

- **Data Science for Everyone**: An absolute zero entry course with no pre-requisites, aimed at getting basic data science concepts out into the general student population beyond Data Science majors. It is a more accessible version of the Berkeley Data 8 class with policy and critical concepts. Perhaps a year-long sequence of courses, divided along GE components?
- **Data Studies Ethics & Policy**: Fulfills Society and Culture breadth requirement and undergraduate writing requirement. Aimed at Data Science majors, but anyone can take this course (no pre-requisites).
- **Data Visualization and Communications**: Fulfills the Arts/Humanities GE requirement
- **Database Management for Data Science**: Focuses on relational database design, data cleaning, no SQL database
- **Intro to Text Mining** or other DH/IR data concepts

Much of this could be drawn from re-adapted current IS UG courses and the new Literacy UC minor.

A3: Professional masters

- A separate program from MLIS, but building on elements from the MLIS
- A new professional degree

A4: Summer institutes

- Focused on techniques and discrete topics for alumni from our various programs
- California Rare Book School (CalRBS) specialization or certificate in Data Ethics? Data Management and Ethics? Data, Society, and Justice? Critical Data Studies?

[SHOULD BELOW COURSES BE MOVED TO RESPECTIVE SECTIONS ABOVE?]

GE courses currently taught within Information Studies:

- **IS 10**: Information and Power
- **IS 20**: Digital Cultures

- **IS 30**: Internet and Society
- **IS 11**: Introduction to Digital Methods: A basic introduction to fundamentals of digital scholarship—data modeling to web presentation* (to be adapted)

Undergraduate minor in Media and Information Literacy:

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

Undergraduate major in Data Studies

INFORMATION ON THIS???

Graduate courses

General

- **IS 289**: Leveraging Big Data of the Past to Address Contemporary Research*
- **IS 275**: Cultural Info/Multimedia - *works with cultural datasets, visualization, text-mining, and other heuristic algorithms, cultural metaphors, interfaces, etc.*
- **IS 289**: AI and Society - *explores data/society issues on economic, political, and design-oriented levels*
- **IS 291C**: Global Media and Information - *explores data's economic value, UBI, data micropayments, platform collectives, data's impact on global movements/social movements, cultural ontologies*
- **IS 272**: HCI Human-Computer Interaction - *teaches basic design, including UX design, information architecture, interaction design*

Professional

- 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

Places where we work [WHAT IS THIS LIST REFERRING TO?]

- JPL
- LAPL
- LAUSD, partner schools, Lab School
- SAADA, etc.
- CRESST partners: Evaluation, assessment, and measurement, particularly in the areas of psychometrics and education

Appendix B: Research, Centers, Partnerships, and Outreach

ITEMS TO ALLOCATE SOMWHERE [BUT DELETE IF DUPLICATE]

- Community-based archives
- POI or C2I2 or DH
- Ozan: https://www.nytimes.com/interactive/2018/04/13/opinion/college-recruitment-rich-white.html
- Borgman stuff

Institutes [NOT SURE WHAT THESE ARE]

- Digital project management
- Data ethics for professionals
- Data design and modelling for scholarly research projects

- Digital curation
- Emulation and digital workstations

B1: DATA for DEMOCRACY

UCLA Data for Democracy in LA 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 comes 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.

B2: CENTER X and Department of Statistics

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 Vision: 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.

The Mission: 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.

Impact: 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.

The Present: 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.

The Immediate Future: 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 that are 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.