ILLINOIS

SCHOOL OF INFORMATION SCIENCES
501 E. Daniel St., MC-493
Champaign, IL 61820-6211

Data Science at the iSchool, University of Illinois Urbana-Champaign

Data Science at the University of Illinois is highly distributed and deeply embedded within nearly all academic units on campus along with Centers such as the National Center for Supercomputing Applications (NCSA) and the recently formed C3.ai Digital Transformation Institute, and in externally facing units such as the Discovery Partners Institute (DPI) in Chicago.

Within the iSchool, data science has always been connected to impact and is envisioned as being tightly integrated into an organizations' existing ecosystem of data, tools, and people. This broader perspective is due in part because the quality of any model depends on the quality of the underlying data on which the model is based, and the impact of a model ultimately depends on users in an organization adopting recommendations made by the model. Most importantly embedding data analytics within the organizational setting ensures that students understand how their analytics activities should be adjusted for the way in which data is collected and are well-poised to identify new data collection strategies that are needed to address a pressing decision. The practices and challenges reported here focus on the Master of Science in Information Management (MSIM) program that was approved by the Illinois Educational Senate in 2015, but work on developing materials for this degree started years earlier in a grant entitled Socio-technical Data Analytics (SODA) and the program goals have already evolved despite being an infant in an iSchool that also houses a #1 ranked program that is over a hundred years old.

Phase I - Socio-technical Data Analytics (SODA) (2011-2016)

A proposal was developed and subsequently awarded from the Institute for Museum and Library Services (IMLS, RE-05-12-0054-12; July 2012 – June 2016) to develop a specialization in Sociotechnical Data Analytics (SODA) at both the master's and doctoral level. Our goal was to fill the workforce gap of people who can analyze data in light of the processes that were in play during data collection and who can effectively communicate the analytical results back to stake holders. The strength of the proposal was in the strategic partnerships with local researchers and businesses who already worked with large datasets. Those alliances enabled MS graduates to receive first-hand experience with both the social and technical implications of large digital data collections, and thus be well-prepared for leadership roles in academic and corporate environments. Similarly, doctoral students could consider multiple stages of the information lifecycle, which ensured that their research findings generalized to a range of scholarly and business settings.

One of the key outcomes the project was the SODA specialization that included both coursework and a structured field experience. Three courses were developed, Evidence-based Discovery (Inaugural offering in Fall 2013 and included in the list of instructors ranked as excellent in Fall 2014), Foundations in Socio-technical Data Analytics (Inaugural offering Spring 2013), and the Socio-technical Data Analytics Practicum (Inaugural offering Spring 2013) which coordinated experiences with industry and academic partners.

At the time the final report was produced for the grant, 19 students had completed the program, 107 MS and 10 PhD students were working towards the specialization and an evaluation of the program had been conducted (Lucic & Blake, 2015). Projects from the project contributed to 21 presentations, 12 posters, 7 conference papers, 3 journal articles, a book chapter and a workshop, many of which are available from the project website http://soda.lis.illinois.edu.

Phase II – Master of Science in Information Management (2016 – current)

As results from the specialization were being distributed, paperwork to form a new Master of Science in Information management was being prepared and the degree, which was approved 2016, launched with two tracks - Data Science and Analytics, and Knowledge Management and Information Consulting. Two of the 3 required courses in the new program leveraged material developed in SODA and the overall program is organized around 4 key pillars: knowledge representation; management and policy; human-centered design and systems; and data analytics. Applications to the program more than doubled from 2019 to 2020 (482 to 985) and acceptance rates also increased as the aspirations of students who applied for the program became better aligned with the program goals, which are:

1. Manage information using best practices in management and policy; knowledge representation; human-centered design and systems; and data analytics.
2. Define and successfully address a tractable research question or real-world problem in information management using the appropriate scientific and/or research methods.
3. Accurately convey the implications of analytical results (in both oral and written modalities) to diverse stakeholders.
4. Articulate the range of values and of ethical standpoints within which complex sociotechnical design occurs.
5. Apply best practices for providing value, leadership, and team building.
6. Stay up to date by learning how to read, analyze, discuss, synthesize, and critique advances reported in the research literature.

Despite the increased number of quality students, the program has faced challenges with rapid changes in policies surrounding international students, and the uncertainty surrounding how courses should be run post COVID-19. The iSchool's LIS program was one of the first programs to be offered online (starting in 1997) so the 20 years of institutional experience and underlying infrastructure made moving courses online less difficult than in some institutions, but many of the community building activities that were conducted in person and needed to be transitioned.

Phase III – The Future (2020 – onward)

Mike Pazzani once lamented that Knowledge Discovery in Databases (KDD) should be called KDF - Knowledge Discovery in Flat files - as the data used in KDD process is rarely stored in a database and is more often stored as a flat file. I would go one step further and say that the potential of Big Data and the accompanying work in Data Science will only be fully realized when model construction is not divorced from the organizational setting in which data is drawn, but rather is an integral part of the socio-technical ecology that continues to evolve with new data sources, methods, and interpretations. Efforts that better define the Findability, Accessibility, Interoperability, and Reuse (FAIR) of digital assets and account for Algorithm Bias also emphasize the importance of the data selection and subsequent decision making that take place after model building. Similarly, efforts such as the Midwest Big Data Hub that combine domain and computational expertise to address key challenges facing society serve as a catalyst to transition Data Science work that is needed throughout the information lifecycle into the into the hands of decision makers in all sectors.

Catherine Blake, PhD
Associate Professor, Program Director (MSIM) and PI, Midwest Big Data Hub (https://midwestbigdatahub.org)
Data Science Leadership Summit, Oct 2020, Challenges and Best Practices for Data Science Institutes & Programs.

Work Cited: Lucic, A. and Blake, C. (2016) Preparing a Workforce to Effectively Re-Use Data, Annual Meeting of the Association for Information Science & Technology, Copenhagen, Denmark, Oct 14-18, 2016.