

Data Science Research at Illinois

What is data science?

If you ask a gathering of scholars and researchers about data science, you are likely to hear as many different definitions of the term as there are people present. In the data science panel held during the Center for Advanced Study Spring Symposium on April 4, 2016, Illinois Astronomy Professor Robert Brunner cited the definition formulated by Allen Renear, dean of Illinois' School of Library and Information Science:

Data science is concerned with all aspects of the creation, management, and analysis of data, focusing in particular on the applications of computational methods to digital data.

This definition expresses a key distinction—that data science is doing things of data rather than with data.

Another way to begin to understand data science is to consider the activities of the many scholars and researchers across the University of Illinois campus, in virtually every discipline, who are doing data science (perhaps without putting that name to their work).

Yet another way to consider data science is as an in-demand skillset and mindset. In 2012, The Harvard Business Review called data scientist “the sexiest job of the 21st century,”¹ declaring that people “with the training and curiosity to make discoveries” have become essential as companies grapple with the challenge of wringing insights from huge volumes and new types of data. “Demand has raced ahead of supply,” The HBR lamented, and “the shortage of

Highlights from the 2016 Center for Advanced Study Spring Symposium.

PANELISTS

Robert Brunner, *Astronomy*

Mara R. Wade, *Germanic Languages & Literature*

Heidi Imker, *University Library*

Matthew Turk, *National Center for Supercomputing Applications*

Ted Underwood, *English*

EDITOR, Trish Barker

data scientists is becoming a serious constraint in some sectors.” As an institution committed to preparing students to meet 21st century challenges and opportunities, the University of Illinois must address this demand for data science.

The state of data science at Illinois

Researchers and scholars across the campus are doing data science, yet these pockets of activity are mostly isolated. While faculty members in fields as disparate as astronomy and early modern studies may be facing similar data challenges, there currently is no effective mechanism for them to learn about one another's activities and to discover opportunities for interdisciplinary collaboration. For campus researchers who would like to acquire data science skills, there is no clear path to follow.

¹ <https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/>

Data Science Research at Illinois

The need for data science support occurs across disciplines. As Heidi Imker, director of the Research Data Service, said:

“People who you might think aren’t interested, are interested. And people who you might think have answers and solutions to all of their data needs already, don’t. What we are seeing is a new need for a broad set of data skills in addition to traditional disciplinary knowledge.”

Mara Wade, professor of Germanic Languages and Literatures, concurred, saying:

“There are many faculty members and graduate students who want to do more data science, but there’s no place to go for basic help and understanding, and that’s really important.”

Undergraduate and graduate students who would like to prepare themselves for data science careers may also be confused about how and where to learn the skills they need. There are opportunities—the Software Carpentry workshops offered by the Computational Science and Engineering (CSE) program,² data management training offered by the Library,³ the informatics minor offered by the Illinois Informatics Institute (I3),⁴ the new MOOC-based data science master’s degree offered by the Department of Computer Science (CS) in collaboration with Coursera.⁵ Unfortunately, because these efforts are evolving independently across campus with little to no coordination and communication, students are likely to be either unaware of these offerings or confused about which options best fit their goals and plans of study. Students of biology, architecture, and English

may want training in data science but may not feel that CSE, I3, or CS is “for them.”

Data science is taking place at Illinois, but the gradual, organic, and isolated evolution of services, education, and training to support data science presents tremendous risks. As our peer institutions move to strategically meet data science needs,⁶ we risk being left behind—and of leaving both our students and faculty underprepared for the educational, career, and research opportunities offered by data science.

Opportunities for Illinois

Data science presents Illinois with enormous opportunities in undergraduate and graduate education and research. Because data science is already generating and evolving in many places on campus, “There are,” as NCSA research astronomer Matt Turk said, “a number of initiatives on campus that can serve as the tinder, that are just looking for a spark.”

Because data science is applicable to virtually all disciplines, but bound by none of them, it offers a virtually unprecedented opportunity to bring together Illinois’ diverse community of researchers and scholars. As Professor Brunner said:

“I don’t know of anything else that can bring people from across campus together when we may know very little about one another’s domains, but data science draws us together.”

As scholars gather at the data science intersection, they will discover novel connections between their research

² <http://cse.illinois.edu/>

³ http://www.library.illinois.edu/sc/services/data_management/

⁴ <https://www.informatics.illinois.edu/informatics-minor/633-2/>

⁵ <https://cs.illinois.edu/news/mooc-based-ms-data-science>

⁶ Notably the Michigan Institute for Data Science (<http://midas.umich.edu/>), Data Science @ Berkeley (<https://datascience.berkeley.edu/>), and the Stanford Data Science Initiative (<https://dsdi.stanford.edu/>)



Data Science Research at Illinois

questions and methodologies, opening up new avenues for innovation discovery.

These novel connections also present opportunities to transform general education by coordinating the different aspects of undergraduate students' education, the qualitative and quantitative intersecting and

“This is a rare opportunity for us to move in a direction that aligns economic significance with research opportunities and opportunities to serve our students through general education.”

informing one another. “In the 21st century, this is a necessity,” said English Professor Ted Underwood. For example, at the University of California, Berkeley, undergraduate students taking the core data science course⁷ are also encouraged to take one of six discipline-specific “connector” courses, ranging from humanities to natural sciences. While they are learning

the foundations of data science, they also are exposed to its relevance to these diverse fields.

“We can’t simply have data science taught in one place,” said Professor Brunner. “It needs to be integrated across the curriculum.”

Professor Underwood emphasized that giving data science a significant place in the undergraduate curriculum is not only educationally beneficial, but would also prepare students for data-science careers, meeting the growing need in industry.

“This is a rare opportunity for us to move in a direction that aligns economic significance with research opportunities and opportunities to serve our students through general education.”

Professor Turk also pointed out that given Illinois’ strengths in developing software, tools, and techniques with broad impact, there is an opportunity to lead in the development of a unified mechanism for interdisciplinary collaboration and the computational tools used for data science.



⁷ <http://www.ischool.berkeley.edu/courses/i8>