INTEGRATIVE DATA SCIENCE INITIATIVE

Robust interdisciplinary collaboration to harness data for the greater good

Sunil Prabhakar
Director of IDSI, Professor of Computer Science
Motivation

“Data are becoming the new raw material of business.”
- Craig Mundie, former head, Res. & Strategy, Microsoft

“The talent shortfall will not only involve data scientists, but it also will extend to existing job classifications from the C-suite to frontlines—all of which are increasingly enabled by analytics.”
- Business-Higher Education Forum

“The diffusion of digital technologies into nearly every business and workplace and pocket has been remaking the US economy and the world of work.”
- Brookings Institute 11/2017
Purdue Data Science Vision

**Interdisciplinary Approach to Research**
- Support structured research efforts in Data Science Theory and Fundamentals, Data-Driven Discovery, and Data Science Applications.

**Pervasive Inclusion of Data Science in Education**
- Establish a *Data Science Education Ecosystem* incorporating data science across campus. Goal: every undergraduate complete her or his studies with relevant professional skills in data science.
- Create physical presences for IDSI and the Educational Ecosystem to promote creative collaboration through proximity and physical interaction.

**Corporate & Non-Profit Engagement**
- Increase data science research and education collaborations with business and industry.

**Foster Entrepreneurship**
ISDI Vision and Launch

Broad engagement of Purdue community
Just one year ago, Purdue University convened town halls and created working groups in both research and education to advance the data science across campus.

“IDSI adds another major goal to the university-wide Purdue Moves initiative.”

180 faculty and staff attended the forums and contributed significant input regarding research, application and educational directions.
Investing in Data Science

Call for Proposals to advance the Data Science Education Ecosystem

- 42 multi-disciplinary proposals were received.
- 9 projects were selected.
- The funded projects include 37 faculty from 22 departments in 10 colleges.

Call for Research Proposals In Data Science

- 52 teams, 172 faculty from 10 colleges and 48 departments submitted.
- 8 were selected for funding.

Related Releases:

Office of the Provost announces request for proposals to build integrative data science education ecosystem

Discovery Park announces request for proposals to support data science initiative

Feedback sought for data science initiative

Follow-up on Purdue’s data science initiative

Faculty, staff input sought to explore new data science initiative
The goal of Data Mind is for every Purdue engineer, starting with the Class of 2020, to graduate with exposure to at least some of the three elements of Data Science: Foundations, Implementation, and Applications.

Integrative Data Science Education Ecosystem

- Prepares Purdue students to invent, innovate, and lead in a data driven world
- Seeks to be highly diverse and multi-disciplinary
- Integrates with our research mission and external partnerships
  - Curricular and lab activities
  - Learning communities
  - Undergraduate research
  - Guest speakers
  - Extracurricular opportunities

College of Engineering

Data Mind

The goal of Data Mind is for every Purdue engineer, starting with the Class of 2020, to graduate with exposure to at least some of the three elements of Data Science: Foundations, Implementation, and Applications.
The Data Mine: A Data Science Living and Learning Community

- Students will live within residential learning communities that enable them to gain data science expertise, data-driven domain knowledge, and participate in early research experiences.
- Perhaps the largest data science Living Learning Community in the nation, creating opportunities for multi-disciplinary undergraduates to learn, grow and thrive together.

Vision: data science for all … infuse into each major; grow data science talent
The Data Mine: A Data Science Living and Learning Community

This fall, 100 students in pilot, scale to 800 students for fall semester 2019.

• Cohorts confirmed for AY 2019-20 include: Critical Data Studies / Libraries / Computer Science / Honors / Anthropology / Actuarial Science / Philosophy / RCHE / VIP (mostly in ECE) / Chemistry / Krannert School of Management / Agriculture

• 50% of the spaces in The Data Mine are reserved for women.

• The STAT-LLC (on which The Data Mine model is based) has 71.4% women this fall, and 4 out of our 21 students this fall are African American women.
Building Data Literacy Skills across Purdue’s Campus

We propose building data literacy among undergraduates by 1) interviewing data science experts on campus and in industry to collect real-life data science stories; 2) using those stories to develop assignments to be integrated into Professional and Technical writing courses; 3) developing both traditional and online course modules that leverage the data science stories; 4) holding pedagogical boot camps to facilitate ethical and persuasive data-rich writing.

Team
Jennifer Bay, Samantha Blackmon, Michael Salvo, Patricia Sullivan, English
Seeding Research

Interdisciplinary Research Projects Initiated

Causally-Driven Healthcare Science – From Observational and Experimental Studies to Personalized and Improved Patient Outcomes
PI: Elias Bareinboim, Assistant Professor, Department of Computer Science, College of Science

Engineering Data Science Algorithms
PI: David Gleich, Associate Professor, Department of Computer Science, College of Science

Fingerprints of the Human Brain: A Data-Science Perspective
PI: Joaquin Goni: Assistant Professor, School of Industrial Engineering and Biomedical Engineering

Quantum Machine Learning for Data Analytics and Optimization
PI: Sabre Kais: Professor, Department of Chemistry

A Relational Based Measure of State Legislator Consequence
PI: Eric Waltenburg: Professor, Department of Political Science
**Supporting Data Science**

**Infrastructure and Community**

IDS will help support research and education in Data-Science
- Hardware & Software
- Personnel
- Data Science Consulting Service

**Building Community**
- Workshops
- Training
- Seminars

PI: Bruno Ribeiro, Computer Science
Leveraging Purdue’s Strengths

**IDSI Collaborates Across Campus**

IDSI works with faculty and centers across Purdue to harness the opportunities at the intersection of domains and data science.

- Center for Science of Information, NSF Science and Technology Center,
- CERIAS
- Joint Transportation Research Program
- Wabash Heartland Innovation Network
- Business Intelligence and Analytics Center
- Purdue Cancer Center
- Regenstrief Center for Healthcare Engineering
- Institute for Global Security and Defense
- …
Upskilling/Reskilling

The Challenge – The number of students majoring/studying in IT fields will not meet the demand for the foreseeable future. We must provide training and educational opportunities to employees educated in more traditional fields to acquire the needed skills in cyber and data analytics as their jobs evolve.

Purdue’s Strategy – Build a campus-wide data science education for all, so that returning and current students learn how all sectors of the economy will be transformed by the internet of things and big data.

Elements of Purdue’s Solution for Companies

• Workshops
• Custom Bootcamps (e.g. Cybersecurity)
• M.S. degrees in Cyber and Data specialties for Engineers and Scientists
• Purdue Global (on-line education)
• Design Thinking for Data Science (early stage)
Partnerships are Key

“The data science initiative is our next major campus wide initiative. Corporations foresee a serious shortfall of talent and seek strategies for company-wide integration of their data assets. Those who come alongside us as Founding Partners will be best positioned to access the talent and innovations of Purdue’s students and faculty.”

- Compelling value for both industry and Purdue
- Top-level commitments
- Multi-year partnership
- Joint oversight/governance
- Joint research planning
- Student engagement
- Campus presence
- Philanthropy

President Mitch Daniels initiated the Purdue Moves agenda, with private-sector partnerships critical to success.

2017-18 Research
$ All-time Records
- $138.2M (30%) private sector
- $65.3M for-profit corps

Integrative Data Science Initiative
Convergence Center for Innovation and Collaboration
Data Science Corporate Council Floor (~ 30k ASF)
~10 Council members co-located, with both private & common space, plus student and faculty teams

SPACES (Regus) Activated Co-Working Space
Professionally managed work space providing flexible options to a variety of tenants

Other Corporate Tenants on 2 Floors

PURDUE Front Door
Focal point for corporate engagement and innovation activities
Council Members play a critical role shaping the future of data science

Founding partners will have differentiating influence on activities and outcomes across data science elements critical to its sector. They will have the opportunity to co-define the direction of the Initiative and will assign commensurate resources that reflect the scale and ambition of Purdue University’s commitment and investment.

- Support and gain preferential access to Purdue’s large and growing interdisciplinary talent-base in the data sciences;
- Inform and influence the Purdue Data Science Initiative;
- Develop unique solutions via convergence of disciplines;
- Advocate for impact for their industry sector;
- Leverage support and knowledge gained from other industry sectors;
- Access new start-ups developed through the initiative.
Thank you, and

Questions?
https://www.purdue.edu/data-science