PURDUE UNIVERSITY | College of Engineering

Engineering Faculty Document
No.: 08-26
April 4, 2025

TO: The Engineering Faculty

FROM: The Faculty of the Division of Environmental and Ecological Engineering

RE: New Engineering Concentration

The Faculty of the Division of Environmental and Ecological Engineering has approved the following new Concentration from the College of Engineering. This action is now submitted to the Engineering Faculty with a recommendation for approval.

TITLE:

Computational Data Science (CDS)

DESCRIPTION:

The focus of the Computational Data Science (CDS) concentration is to educate students on topics related to data science and computation, and to enable proficiency in the use of state-of-the-art tools and techniques related to the field of computational data science.

Students completing the CDS concentration will be able to:

- Utilize data and computation to solve challenging problems in their field of study.
- Communicate and collaborate with individuals from diverse interdisciplinary backgrounds.
- Obtain, organize and manipulate large datasets in their field of study.
- Dynamically analyze and visualize novel data relationships.
- Conduct appropriate statistical analyses for large datasets in their field of study.

RATIONALE:

With the increased demand for computational skills in the fast-growing field of data science, the CDS concentration aims to enable students to enhance their graduate education with interdisciplinary skills in computational data science and to prepare them for future challenges in this competitive industry. CDS is offered by the Computational Interdisciplinary Graduate Program (CIGP). Graduates with the CDS concentration should be well prepared to join interdisciplinary teams and become leaders in research and development. While there are some data science-related programs available at Purdue currently, there is no concentration in 'Computational Data Science' that is specifically geared towards graduate students at Purdue. The CDS concentration is an ideal mechanism for graduate students to acquire these skills because it allows flexibility to

pursue another specialization while already being enrolled in their primary major. Please see the attached proposal for details.

Head/Director of the Division of Environmental and Ecological Engineering

Link to Curriculog entry:

EEE MS: Computational Data Science - CDS - -Division of Environmental and Ecological Engineering - WL

EEE PhD: Computational Data Science - CDS - -Division of Environmental and Ecological Engineering - WL

Computational Data Science (CDS) track for
Computational Interdisciplinary Graduate Programs (CIGP)

With increased demand for computational skills in the fast-growing field of data science, the aim of the proposed Computational Data Science (CDS) concentration is to enable students to enhance their graduate education with interdisciplinary skills in computational data science and to prepare them for future challenges in this competitive industry.

While there are existing data science-related programs available at Purdue currently (such as the Integrative Data Science Initiative, The Data Mine, and individual courses offered by different majors), there is no concentration in 'Computational Data Science' that is specifically geared towards graduate students at Purdue. Short of a full-fledged graduate degree in data science, an interdisciplinary concentration in CDS would be an ideal mechanism for graduate students to acquire these skills because students usually do not have the flexibility to pursue another degree program while already being enrolled in their primary major.

The Computational Interdisciplinary Graduate Program (CIGP) is uniquely positioned to address this need for data science education due to its collaborative and interdisciplinary nature. If approved, CIGP will be able to offer the following mutually complementary tracks:

- Computational Science and Engineering (CS&E) with concentrations CMEN and CMSI.
- Computational Life Sciences (CLS) with concentration CMLS.
- Computational Data Science (CDS) with concentration CMDS ***(Proposed)***.

The three tracks will enable students to focus on specific areas of computing that are of interest to them and will likely attract graduate students from various departments across Purdue. This will help cultivate interdisciplinary researchers and scholars who can leverage their computing skills to tackle problems that cut across traditional disciplinary boundaries.

Proposed Requirements for students to complete the CDS Concentration:

Requirements for the CDS concentration will be similar to those for the existing CSE and CLS tracks:

- For MS:
    - Two 'core' courses (6 credits) and one 'relevant' course (3 credits) with a minimum grade of 'B' in each.
    - One semester of CIGP seminar.
    - MS Poster presentation (during final semester).
- For PhD:
    - Two 'core' courses (6 credits) and two 'relevant' courses (6 credits) with a minimum grade of 'B' in each.
    - Two semesters of CIGP seminar.
    - PhD talk (during final semester).

Note that the two 'core' courses must be taken from two different core areas.

Additionally, some departments or courses may have restrictions on which classes students are allowed to take and/or put onto their Plan of Study. Students should consult with their home department when faced with any such restrictions.

Core Areas for CDS: Most listed courses have been taught at least twice in the last 5 years:

- **Data Analytics & Visualization** *(partial overlap with 'Scientific Visualization' core area for CSE)*:
    - CGT 57500 - Data Visualization Tools And Applications
    - CGT 58100 - Medical Image Processing and Visualization
    - CGT 67000 - Applications In Visual Analytics
    - CNIT 58100: Data Analysis
    - CS 53000 - Introduction to Scientific Visualization
    - CS 57300 - Data Mining
    - CS 59300TDA - Topological Data Analysis
    - ECE 66100 - Computer Vision
    - ECET 54900 - Advanced Applied Computer Vision for Sensing and Automation
    - STAT 65600 - Bayesian Data Analysis

- **Intelligent Computing** *(overlap with core area for CSE)*:
    - BME 64600 - Deep Learning
    - CS 57100 - Artificial Intelligence
    - CS 57700 - Natural Language Processing
    - CS 57800 - Statistical Machine Learning
    - CS 58700 - Foundations of Deep Learning
    - CS 59300CVD - Computer Vision with Deep Learning
    - CS 59300MLT - Machine Learning Theory
    - ECE 57000 - Artificial Intelligence
    - ME 53900 - Introduction to Scientific Machine Learning

- **Probability and Statistics:**
    - CNIT 60100 - Applied Statistics In Information Technology
    - MA 51900 - Introduction to Probability
    - MA 53200 - Elements of Stochastic Processes
    - STAT 51200 - Applied Regression Analysis
    - STAT 52500 - Intermediate Statistical Methods
    - STAT 52600 - Advanced Statistical Methods
    - STAT 52900 - Bayesian Applied Decision Theory
    - STAT 54500 - Introduction to Computational Statistics
    - STAT 54600 - Computational Statistics

- **Programming and Computing** *(overlap with core area for CSE and CLS)*:
    - BIOL 59500 - Practical Biocomputing
    - CS 50100 - Computing for Science and Engineering
    - CS 51400 - Numerical Analysis
    - CS 51500 - Numerical Linear Algebra
    - CS 52500 - Parallel Computing
    - IE 54100 - Nature-Inspired Computing
    - MA 57400 - Numerical Optimization
    - STAT 50600 - Statistical Programming And Data Management
    - STAT 52700 - Introduction To Computing For Statistics

- **Data Science Applications:**
    - AAE 59000 - Data Science Applications in Mechanics of Materials
    - ABE 53100 - Instrumentation And Data Acquisition
    - AT 60700 - Aviation Applications Of Bayesian Inference
    - BCHM 61200 - Bioinformatic Analysis Of Genome Scale Data
    - CE 50701 - Geospatial Data Analytics
    - CE 59700 - Image-based Sensing
    - EAPS 50700 - Introduction to Analysis and Computing with Geoscience Data
    - EAPS 51500 - Geodata Science
    - FNR 57400 - Big Data, AI, And Forests
    - HSCI 52500 – Statistics and Computational approaches for Health Sciences
    - TDM 51100 - Corporate Partners

Relevant Courses for CDS:

- ABE 65100 - Environmental Informatics
- AGRY 54500 - Remote Sensing Of Land Resources
- AGRY 64100 - Statistical Hydrology
- AT 50700 - Quantitative Research Methodologies In Transportation
- BIOL 56310 - Protein Bioinformatics
- BIOL 58210 - Ecological Statistics
- BIOL 59500 - Introduction to Bioinformatics
- BIOL 59500 - Practical Biocomputing
- BME 50100 - Multivariate Analyses In Biostatistics
- CE 50801 - Geographic Information Systems
- CE 61400 - Statistical And Econometric Methods I
- CEM 53300 - Infrastructure Analytics
- CNIT 57000 - IT Data Analytics
- CNIT 58100: Data Literacy

continued over ..

- CS 50023 - Data Engineering I
- CS 50024 - Data Engineering II
- CS 52000 - Computational Methods in Optimization
- CS 52900 - Security Analytics
- CS 53600 - Data Communication And Computer Networks
- CS 54100 - Database Systems
- CS 54200 - Distributed Database Systems
- CS 55600 - Data Security And Privacy
- CS 57900 - Bioinformatics Algorithms
- CS 59200BDS - New Trends in Big Data Systems
- CS 59200DOM - Distributed Optimization for Deep Learning
- CS 59200MLS - Machine Learning Systems
- CS 59200RLI - Reinforcement Learning
- ECE 56200 - Introduction To Data Management
- ENGT 58300 - Applied Engineering Statistics For Industry
- FNR 55800 - Remote Sensing Analysis And Applications
- FNR 58700 - Advanced Spatial Ecology And GIS
- FNR 64700 - Quantitative Methods For Ecologists
- IE 53300 - Industrial Applications Of Statistics
- IE 53600 - Stochastic Models In Operations Research I
- MA 51100 - Linear Algebra with Applications
- PSY 60601 - ANOVA For The Behavioral Sciences
- PSY 60800 - Measurement Theory And The Interpretation Of Data
- PSY 63100 - Multiple Regression Analysis For The Behavioral Sciences
- STAT 51300 - Statistical Quality Control
- STAT 52000 - Time Series and Applications
- STAT 52200 - Sampling and Survey Techniques
- STAT 52400 - Applied Multivariate Analysis
- STAT 58200 - Statistical Consulting and Collaboration
- TDM 50100 - The Data Mine Seminar
- TDM 59000 - The Data Mine Special Topics
- TECH 62800 - Technology Research And Use Of Data Analytics

Questions / Feedback on the draft CDS proposal circulated in Spring 2024 (and responses):

- **Is there a need for the proposed CDS concentration? Is CIGP the appropriate home for CDS? Consider consulting the Bureau of Labor Statistics (BLS) (www.bls.gov) for career trends and also the O*NET (Occupational Network) database (www.onetonline.org) that is built from existing job descriptions so the core competencies are identified.**

    - Indeed, The Bureau of Labor Statistics lists data science among the fastest-growing occupations. The bureau projects that the field will keep growing by more than 30% in the coming decade: https://www.bls.gov/ooh/math/data-scientists.htm
    - The Occupation Network also provides information on related skills: https://www.onetonline.org/link/summary/15-2051.00 listing Python, R, etc. as 'in-demand' skills.
    - See further justification on Page 1.

- **How is the proposed CDS concentration different from CLS and CSE? Could one augment existing CLS and CSE concentrations with data-related courses, and change their name, for instance to Computational and Data Science for Life Sciences. It needs to be more clearly differentiated and the difference in the target student groups better delineated. More detail on the target student group(s) would make a more compelling case.**

    - CSE and CLS have traditionally been aligned with scientific computation with more focus on computation and less on data. On the other hand, Data Science has been centered more around data analytics with lesser focus on the "physics" behind it.
    - Currently, we see several CIGP students gravitating towards courses in the 'Intelligent Computing' core area in CSE-CIGP – a catch all for data science-related courses. Given the recent upsurge of AI/ML, this is not surprising. However, these CIGP students are forced to take either the CSE or CLS option, whereas the proposed CDS would be more appropriate for these students because scientific computing/simulations is not their primary interest.
    - The proposed CDS concentration will also help differentiate those CSE/CLS students who are engaged in traditional computational science (scientific computing/simulations) from students in CDS who are more interested in data-related topics. For students who are engaged in both traditional CSE/CLS tracks and in data science, the proposed CDS concentration will allow them to choose their primary interest.
    - There will inevitably be some overlap between all the 3 CIGP tracks (CSE/CLS/CDS), but there are also core areas that are exclusive to each that differentiate them from each other.

- **On "Examples of existing graduate programs related to Data Science at peer institutions in the US:" I realize these programs are diverse, but it would be good to provide a summary of required student background and number/type of courses required.**

    - There are several graduate-level programs in Data Science across the US catering to different students - some online programs focusing on working professionals and some for on-campus students. Depending upon the nature and focus of the program, they also differ in the

number/type of courses required. For instance, there are programs ranging from minor specializations to full-fledged degrees. Programs offered by Management departments are more focused on business analytics in comparison to programs offered by science and engineering departments.
    - Our focus in CIGP would be on developing a "concentration" that would be available to any graduate student from a department or school affiliated with CIGP.

- **On course requirements, "Applications" should be included as a Core area since data science is nothing if there is no data.**
    - Indeed, "Applications" is a proposed Core area for CDS.

- **Will all departments and schools at Purdue be eligible to participate?**
    - Initially, interest in participation is being sought only from current CIGP affiliated departments. If and when CDS is approved and instituted, all departments and schools will be welcome to be affiliated with CIGP and participate in CDS/CSE/CLS tracks pending departmental approval.

- **What effect will CDS have on CIGP enrollments (and in particular on CSE and CLS enrollments)?**
    - Anticipate that overall CIGP enrollment may increase.
    - However, CSE and CLS enrollments may decrease because currently, students interested in data and computing can only choose between CSE or CLS.

- **Will Faculty be able to participate in multiple CIGP tracks (CSE/CLS/CDS)? What value does CDS (and in general, CIGP) offer to participating faculty?**
    - Yes.
    - CIGP faculty can nominate and recruit students using Lynn Fellowships.
    - Enhance visibility of their interdisciplinary teaching and research related to computing.
    - Foster collaborations.

- **Will CIGP need to have faculty representatives from each of the CIGP tracks (CSE/CLS/CDS)?**
    - Yes. If an affiliated department or school participates in multiple CIGP tracks, there will need to be a faculty representative for each track.
    - It is up to the department or school to decide whether the same faculty member may serve as the representative for multiple CIGP tracks or if different faculty members will serve as representatives for different tracks.

- **Next steps for CDS:**
    - CIGP faculty representatives vote on the proposal for CDS concentration.

- Invite CIGP affiliated departments (through their Representatives) to participate in CDS and seek informal approval from their Heads and/or Graduate Committees as needed.
- Submit to Curriculog for formal approval – see Appendix.

Existing programs at Purdue related to Data Science:

- Integrative Data Science Initiative (IDSI): https://www.science.purdue.edu/data-science/
  - BS major in Data Science offered by CS & STAT (for UG)
  - Certificate offered by Integrative Data Science Initiative (for UG)
  - Online modules (for Graduate Students): 1-credit 5-week modules
- The Data Mine (TDM): https://datamine.purdue.edu/
  - Residential Learning Community – Primarily UG, but Graduate students too
  - Project-based courses with corporate partners
- Purdue Engineering Online Certification (Professional Education): https://engineering.purdue.edu/online/certifications/data-science
- Online SimpliLearn Graduate program offered by Purdue & IBM: https://www.simplilearn.com/pgp-data-science-certification-bootcamp-program
- MS in Business Analytics and Information Management (MSBAIM): https://programs.business.purdue.edu/msbaim/domestic-plan-of-study/
- Master's in Applied Geospatial Analytics - College of Agriculture, College of Liberal Arts, and Polytechnic Institute at Purdue https://polytechnic.purdue.edu/degrees/masters-applied-geospatial-analytics-online/applied-geospatial-analytics-plan-of-study
- Several UG/Grad Courses at Purdue related to data science.

Examples of existing graduate programs related to Data Science at peer institutions in the US:

- MIT: Professional Education Certificate: https://professional-education-gl.mit.edu/mit-applied-data-science-course
  - 12-week program geared towards professionals.
  - Project-based as opposed to credit-based.

- Stanford: MS in Statistics and Data Science: https://statistics.stanford.edu/graduate-programs/statistics-ms/statistics-data-science-curriculum
  - Students typically finish in 5-6 quarters (not including summer).

- Northwestern: MS in Data Science: https://sps.northwestern.edu/information/data-science-online-masters.php

- Columbia University: MS in Data Science (MSDS) https://datascience.columbia.edu/education/programs/m-s-in-data-science/

- University of California, Berkeley: Master of Information and Data Science (MIDS) https://ischoolonline.berkeley.edu/data-science/

- University of Michigan: MS in Data Science
  https://cse.engin.umich.edu/academics/graduate/graduate-programs/masters-in-data-science/

- Harvard University: MS in Data Science
  https://seas.harvard.edu/applied-computation/graduate-programs/masters-data-science

- Georgia Institute of Technology: Master of Science in Analytics
  https://info.pe.gatech.edu/oms-analytics/

- University of Texas at Austin: Online Master of Science in Data Analytics and Information Systems
  https://cdso.utexas.edu/msds

- Carnegie Mellon University: MS in Computational Data Science
  https://mcds.cs.cmu.edu/apply-mcds-program

and several others …

Appendix:
Information needed for Curriculog: Computational Data Science

Program Type – Always Select Program: Program

Level and Campus: Graduate-PWL

PWL Only: Will this be offered in Indianapolis?: No/Not Applicable

College/School: TBD

Department: TBD

New Concentration Name: Computational Data Science

Program(s) and Major(s) for which concentration applies (e.g., BIOL-BS; BIOL-MS): TBD, but will include program(s), major(s), and concentration.

Will new courses be created for this concentration?: No

PWL – Is the concentration managed by OIGP?: Yes

Total Credits: 12

How does this concentration align and support the host program(s) and major(s)? Is it Optional or Required for the host program(s)? Address distinction from other available concentration(s).

With the increased demand for computational skills in the fast-growing field of data science, the aim of the proposed Computational Data Science (CDS) concentration is to enable students to enhance their graduate education with interdisciplinary skills in computational data science and to prepare them for future challenges in this competitive industry. CDS would exist within the Computational Interdisciplinary Graduate Program (CIGP), which is an optional concentration with the aim of producing MS and PhD students who have learned about computational tools and techniques in one or more areas of computational science. Graduates with concentrations in CIGP should be well prepared to join and make significant contributions to Interdisciplinary research teams and are expected to become leaders in research and development.

While there are some data science-related programs available at Purdue currently (such as the Integrative Data Science Initiative, The Data Mine, and individual courses offered by different majors), there is no concentration in 'Computational Data Science' that is specifically geared towards graduate students at Purdue. Short of a full-fledged graduate degree in data science, an interdisciplinary concentration in CDS would be an ideal mechanism for graduate students to acquire these skills because

students usually do not have the flexibility to pursue another degree program while already being enrolled in their primary major.

The program is designed to produce MS and PhD graduates who will have the knowledge and skills to utilize data and computing to solve problems in their major field of study and beyond. Graduates with the Computational Data Science (CDS) concentration should be well prepared to collaborate on interdisciplinary research teams and make significant contributions to research, development and practice. The **Department Name(s)** will be the first of many partnering home departments.

Summarize the benefits to the target audience.

Unique knowledge or abilities of students with this concentration are in computational tools and techniques in science and engineering. In order to graduate with the CDS concentration, students will need to have taken a required combination of core courses and relevant courses from various disciplines (see attached), as well as attend the required number of CIGP seminars. The number of core and relevant courses a student must take depends on the type of degree the student is pursuing. Master's students will need to complete two core courses, one relevant course, and one seminar course for a total of 9 credit hours. PhD students will need to complete two core courses, two relevant courses, and two seminar courses for a total of 12 credit hours. CIGP seminars will be composed of students seeking a concentration in Computational Science and Engineering (CSE), Computational Life Sciences (CLS) and, if approved, Computational Data Science (CDS) to expose students to a variety of interdisciplinary topics related to computation. Students in participating departments with this concentration would have to demonstrate the ability to communicate to peer audiences with a required 3MT during the seminar and by completing a PhD talk or MS Poster in their last semester before graduating.

Describe the research focus and/or career relevance.

The focus of the Computational Data Science (CDS) track is to educate students on topics related to data science and computation and to enable them to become proficient in the use of state-of-the-art tools and techniques related to the field of computational data science.

Students completing the Computational Data Science (CDS) concentration will be able to:

- Utilize data and computation to solve challenging problems in their field of study.
- Communicate and collaborate with individuals from diverse interdisciplinary backgrounds.
- Obtain, organize and manipulate large datasets in their field of study.
- Dynamically analyze and visualize novel data relationships.
- Be familiar with appropriate statistical analyses for large datasets in their field of study.
- Communicate and collaborate across disciplinary boundaries.

Projected Headcount: 25+

Does the concentration call for a new fee that is not already in use?: No (most common).