

Data Science at Rensselaer
James Hendler
Director, Rensselaer Institute for Data Exploration and Applications
RPI

In 2018, Rensselaer Polytechnic Institute (RPI), the nation's oldest technological university, introduced an innovative new requirement as part of its overall curriculum revision. RPI became the first school to require that every undergraduate, regardless of major, would need to take at least two "data intensive" courses to graduate¹.

The motivation for this requirement, as stated in the undergraduate curriculum is that we live in a world that is increasingly data and information-rich and the fields for which we are preparing our students are increasingly data-driven. Data are being collected in all areas of the human experience, and massive collections of this data are now readily available online and elsewhere. In aggregate, these data collections are providing unprecedented opportunities for growth in human understanding. Concomitantly, data analytics and interpretation are emerging as fundamental components of daily life (e.g., social media, online commerce, search engines, web media, infrastructure optimization and management, electronic medicine, etc.). Data analysis and modeling are becoming ubiquitous throughout science, technology, engineering, medicine, the humanities, arts, and social sciences, and business. Students will almost certainly encounter data, data analysis, and the results of data analysis in their careers and daily lives, regardless of their major and career path. As data consumers, they will at a minimum need to understand how to use data interpretation and data analytics for decision making. Many will be creators of data analysis solutions in the technological workforce, and others will play decision-making roles where they will have to leverage data resources, or manage their creation and maintenance.

Rensselaer has taken a leadership position in defining a well-educated student in the new data driven economy and world by making sure that all students, regardless of major, have data awareness and data dexterity. Through its Data Intensive, or DI, requirement, Rensselaer aims to develop graduates who are sophisticated consumers, producers, and users of data. Our graduates will be able to interpret and/or transform data into actionable insight and solutions to compelling problems and then communicate associated findings effectively to diverse audiences. To best prepare our students to operate in this rapidly evolving environment, provide the global workforce with thought leaders, and create breakthroughs that will become tomorrow's new products and solutions, Rensselaer has made data dexterity a core part of the undergraduate experience. Through a university-wide effort, Rensselaer has defined a distinctive integrated interdisciplinary undergraduate requirement in data dexterity. The data dexterity requirement begins with the premise that all Rensselaer students will acquire basic data awareness and literacy skills as part of their core undergraduate experience.

¹ <https://www.chronicle.com/newsletter/teaching/2018-04-05>

To accomplish this goal, the Institute created a sequence of courses with the following “data dexterity” objectives:

- Identify different types of data, information and evidence within the relevant discipline, and be able to discuss issues of data quality, curation, validation, and uncertainty.
- Identify appropriate problems to which data can be applied, and discuss limitations, biases, assumptions and interpretations.
- Determine appropriate analytical tools and effectively use them with relevant data types to formulate, analyze, interpret, and/or solve real-world problems.
- Effectively communicate about problems/issues in this field in which data is a relevant tool, including writing about, presenting on, and visualizing data.
- Discuss the ethical issues surrounding data in this field, including, but not limited to, responsible conduct of research, privacy, provenance, privatization, monetization, and social implications.

As of 2020, this requirement went into effect for all students. Figure 1 shows a sample of the DI-1 (intro) and DI-2 (advanced) courses available for this requirement.

DI-1	ARCH	2360	ENVIRONMENTAL AND ECOLOGICAL SYSTEMS	DI-2	ARCH	2370	Energy Comfort and Ecology
DI-1	ARCH	4090	Architectural Case Studies	DI-2	BIOL	4630	Molecular Biology II
DI-1	ASTR	2120	Earth and Sky	DI-2	IENV	4700	One Mile of the Hudson
DI-1	BIOL	1010/1015	Intro Bio	DI-2	MGMT	2100	Statistics
DI-1	CHEM	1200	Chem II	DI-2	PHYS	2350	Experimental Physics
DI-1	COMM	2120	Language in Real Time				
DI-1	COMM	2570	Typography	DI-2	COGS/CSCI/PSYC	4210/496x/4510	Cognitive Modeling
DI-1	COMM	2960	2D-Motion Graphics	DI-2	ECON	4570	Econometrics
DI-1	CSCI	1100	CS 1	DI-2	PSYC	4310	Research Methods & Statistics II
DI-1	ENGR	2600	Modeling and Analysis of Uncertainty	DI-2	BIOL/BCBP	4660/6650	The Biology of Systems
DI-1	ERTH	1250	Geo II:lab	DI-2	STSS	4990	STS and Sustainability Senior Project
DI-1	ERTH	2140	Introduction to Geochemistry	DI-2	COGS/CSCI	4410/4966	Programming for COG SCI & AI
DI-1	ERTH	2330	Earth Materials	DI-2	COMM	4690	Interface Design: Theory and Application
DI-1	IHSS	1200	Principles of Economics	DI-2	STSS/H	4970	DSGN & INNOVATION SENIOR PROJ
DI-1	IHSS	1977	War and Technology: Past, Present, Future	DI-2	COMM	4370	Information Design
DI-1	MATH	1520	Math for Management and Economics	DI-2	ARTS	4160	Music and Technology II
DI-1	MGMT	2510	Intro to Data Analytics	DI-2	CHEM	4110	Instrumental Methods of Analysis
DI-1	COGS	2100	Critical Thinking	DI-2	COMM	4880	Interactive Data Visualization
DI-1	STSS	4961	History of Science and Technology	DI-2	BIOL	4200	Biostatistics
DI-1	CSCI	1200	Data Structures	DI-2	ARTS	4130	New Media Theory
DI-1	BIOL	2120	Intro to Cell and Molecular Biology	DI-2	CHME	4150	Chem Engineering Lab I
DI-1	ERTH	1100/1150	Geology I	DI-2	MANE	4510	Control Systems Lab
DI-1	COGS/PSYC	4330	Intro to Cognitive Neuroscience	DI-2	MANE	4020	Thermal & Fluids Eng Lab
DI-1	COGS/PSYC	4360	Behavioral Neuroscience	DI-2	BIOL	4220	Machine Learning for Environmental Biology
DI-1	COGS/PSYC	4610	Stress and the Brain				
DI-1	COMM	2520	Communication Theory and Practice				
DI-1	STSS	2100	Investigating Society				

Figure 1 Examples of Rensselaer's Data Intensive Courses

Acknowledgments

The idea for the data dexterity requirement came to a large degree from an NSF grant to Dr. Kristin Bennett (Data Analytics Throughout Undergraduate Mathematics, DMS award 1331023). Dr. Bennett and Dr. Peter Fox generated the original list of the objectives which were then modified and adopted by the Institute Curriculum Committee and approved by the State of New York Department of Education as a component of Rensselaer's undergraduate requirements update.