Empowering people to better design and tackle data-rich research projects
About the Data Science Hub
Meet the Facilitators and Steering Committee
Our Connections
Coding Meetup
Teaching Tech & Instructors
Community Events: Research Bazaar & ML+X
By the Numbers
Future Goals
Since 2018, the Data Science Hub has been providing data science training and implementation across campus. The Data Science Hub is part of the Wisconsin Institute for Discovery and collaborates closely with the American Family Insurance Data Science Institute.

The Data Science Hub executes a mission for community engagement and learning opportunities for campus researchers through a variety of services:

- **consultations** with Data Science Facilitators who can recommend learning pathways & project strategies, and liaise contacts with collaborators & data science experts

- **community events** and co-sponsored seminars that bring together researchers and other partners around relevant data science topics

- **regular trainings** (including Carpentries workshops) around fundamental data science and computational skills

These activities are led by the Data Science Hub Facilitators, with support and advising from the Data Science Hub Steering Committee.
MEET THE DATA SCIENCE HUB

Chris Endemann
Data Science Facilitator

Clare Michaud
Data Science Facilitator

Sarah Stevens
Data Science Facilitator

Michael Ferris
Data Science Hub Director

Data Science Hub Steering Committee

Laura Albert, Industrial and Systems Engineering
Victor Cabrera, Animal and Dairy Sciences Department
Cameron Cook, Libraries & Research Data Services
Russell Dimond, Social Sciences Computing Cooperative
Michael Ferris, Computer Sciences & Wisconsin Institute for Discovery
Iain McConnell, Data Science Institute & Space Science Engineering Center
Lauren Michael, Center for High Throughput Computing
Michael Newton, Statistics & Biostatistics and Medical Informatics Departments
Amanda Smith, Industrial and Systems Engineering
Whitney Sweeney, Data Science Institute
Brian Yandell, Data Science Institute & Statistics Department
The Data Science Hub’s facilitative work relies on strong connections with other data science and research support professionals at UW-Madison, in the Madison community, and at other institutions. In addition to the Data Science Institute and the Wisconsin Institute for Discovery, here’s a look at groups that the Data Science Hub collaborated or engaged with in 2021:

- UW-Madison Libraries
- Ebling Library & Health Science Learning Center
- Research Data Services
- DoIT’s Research Cyberinfrastructure group
- Department of Applied Agricultural Economics
- Department of Plant Pathology
- Nelson Institute for Environmental Studies
- UW-Madison Office of Sustainability
- Center for High Throughput Computing
- Tiny Earth (Wisconsin Institute for Discovery)
- UW-Madison Carpentries Community
- Midwest Carpentries Community
- Global Carpentries Community
- Marquette University
- University of Wisconsin-Milwaukee
HERE TO HELP: CODING MEETUP

Twice a week, the Data Science Hub Facilitators host coding meetup for anyone on campus to ask questions about using computational tools and methods in their work. The Data Science Hub seeks to foster community around data science through these informal and collaborative meetups: it’s always encouraged for those attending to help one another. Coding meetup broadens the Hub’s reach to create a network of data science support. There were 91 visits to Coding Meetup in 2021, with 32 unique attendees.

“After attending a tech training class, you can find yourself feeling like you aced the coursework and assignments, but then when it comes time to try and apply what you’ve learned on real-world data sets and problems, you run into stumbling blocks you just don’t know how to troubleshoot with the basic skills. You can really feel lost and stuck having no resources to help, and sometimes even give up on the new technology altogether. Office Hours is an antidote to this: a way of extending the training to supporting growth and learning for individuals in their actual practices and fields of work...Office Hours has been instrumental to my continued use of the programs I learned in a Carpentries workshop...Office Hours is what differentiate [the Data Science Hub] from all other technology continuing education opportunities I’ve been a part of.

— Catherine Harris, Pollution Prevention Specialist, Madison Metropolitan Sewer District
Since 2019, the Data Science Hub has been hosting Carpentries workshops for the UW-Madison campus community, carrying on work started by the Advanced Computing Initiative and then the Center for High-Throughput Computing. Hosted by the Data Science Hub, organizing and teaching at these workshops is community-driven, supported by dozens of instructors from across campus, with a range of subject-area expertise.

During 2021, the 15 workshops that the Data Science Hub hosted reached all levels of researchers from departments on-campus and organizations outside of the University. The workshops centered around the subject areas of ecology, health sciences, social sciences, and geospatial science, with the shared goal of showing researchers best practices for data management, data creation, and software development. Workshop topics taught for the first time in 2021 were Introductions to Machine Learning and Interactive Data Visualization.

The Data Science Hub also had the opportunity to teach the broader community: a workshop on GitHub was offered through ForwardFest.

[The workshops] introduced me to tools I use in my day-to-day research and work. My attending the workshops has given me confidence to take my research beyond the methods that are commonly employed in my field. For example, I continued my data collection through the worst of the pandemic thanks to some retooling of Qualtrics.

— Erwin Lares, PhD Candidate, Department of Spanish and Portuguese
In addition to providing trainings in foundational computational skills, the Data Science Hub, as the UW-Madison Carpentries Community’s home, trains new Carpentries instructors in interactive and evidence-based pedagogy to teach researchers how to apply data science methods and tools to their work. This training proves to be useful for people with teaching experience and people new to teaching alike. Through Instructor Training, the local Carpentries Community grows, and the trained instructors are able to bring what they learn into their own educational spaces — be they classrooms, libraries, or one-on-one consultations. The Data Science Hub hosted two instructor trainings in 2021, training 8 new UW-Madison-affiliated instructors.

"The instructor training provided me with fundamental educational background, lesson design strategies, and techniques to use in workshop-style teaching. Even though I am a professional educator, the instructor training was the first time I was taught very basic concepts like different types of assessment strategies, Bloom’s taxonomy, and how to design learning objectives. I learned how to guide learners through a live demonstration of a technical task and how to make the educational environment welcoming to all. The instructor training had numerous impacts on my teaching, both on the formal courses I teach and the Machine Learning for Biology workshop my research group created. When moving CS 540: Introduction to Artificial Intelligence to an online format for the Fall 2020 semester, I used my experience from instructor training to restructure the lecture format. My co-instructors and I introduced extensive formative assessments, for instance, using short quiz questions in between lecture segments to check students' progress and understanding. In addition, I used my improved understanding of learning objectives when working with colleagues in the BMI department to develop a new course proposal for BMI/CS 775: Computational Network Biology, which was approved in Fall 2021."

— Anthony Gitter, Associate Professor, Department of Biostatistics and Medical Informatics
Teaching Resources

Through growing the community of workshop instructors, the local Carpentries Community has been able to reach new audiences of learners and collaborate with instructors at other institutions to create new, relevant teaching material. Work on new instructional materials is underway for **high-dimensional data analysis, deep learning, and text analysis for the humanities**. We hope that offering these new workshop topics broadens that audience that the Data Science Hub reaches and strengthens our collaborative relationships.

Research Resources

Data Science Hub Facilitators led efforts to create the new [Researcher Toolkit](#), in collaboration with colleagues in the Wisconsin Institutes for Discovery, the Libraries, and Research Cyberinfrastructure. The Researcher Toolkit guides researchers through each aspect of the research lifecycle: from devising a research question to gathering, analyzing and visualizing data, to publishing outputs and archiving anything related to the research.

*A visual guide for Research Toolkit users to refer to as they think through planning a research project and use the resource.*
The Data Science Hub hosted its second Data Science Research Bazaar in 2021, with the theme of Data Science for the Social Good. Throughout February, presenters and attendees learned from each other during lightning talks, posters, science-inspired artwork, discussions, and workshops that addressed data science’s intersections with racism, health and the environment, and cities.

The Research Bazaar is an annual event to bring together members of the UW-Madison campus and the surrounding community to exchange ideas and knowledge about data. From hosting this event repeatedly, the Data Science Hub has fostered connections among researchers and data scientists. In 2021, a mix of 350 individuals participated as presenters and attendees, including undergraduate students, graduate students, post-docs, staff and faculty members, industry employees, and community members.
COMMUNITY EVENTS: ML+X

The Data Science Hub has been able to grow in new directions in 2021, including with providing more robust support around machine learning. Through forming the Machine Learning Community of Practice, the Data Science Hub aims to support and connect machine learning practitioners at UW-Madison.

A new regularly occurring event for the campus community that the Data Science Hub and the Machine Learning Community of Practice host is ML+X.

Once a month, the Machine Learning Community of Practice gathers for ML+X’s short presentations, where researchers and data scientists share their projects and use cases for machine learning techniques and tools.

Each month, researchers and data scientists gather on Zoom for ML+X.
BY THE NUMBERS

Drop-in consultations at coding meetup 91

Unique visitors to coding meetup 32

Total workshops hosted (including instructor training) 17

Total workshop attendees 219

Newly trained Carpentries instructors 8

Research Bazaar participants 350
In 2022, the Data Science Hub looks forward to building on previous years’ work to expand its services and reach more members of the UW-Madison campus and the local community. The activities that the Data Science Hub initiates and maintains reflects its Strategic Goals for 2020-2023:

1. **Build data science competency among researchers**
The Data Science Hub hosts at least 15 training workshops each year. In 2022, new workshop topics will include text analysis for the humanities and deep learning, and lesson development.

2. **Foster real-time engagement and collaboration**
The Data Science Hub is hosting its 3rd annual Research Bazaar, continuing to host Coding Meetup twice a week, and has introduced a lightning talk series on machine learning, ML+X.

3. **Connect researchers to resources and collaborators**
The Facilitators reach researchers across campus by presenting at department seminars and new student orientations, and making up-to-date resources for research projects widely available.

4. **Strengthen relationships with resource providers**
The Data Science Hub has maintained relationships with frequent collaborators, and forged new ones: dataiku, Snowflake, and WARF are all new sponsors of the 2022 Research Bazaar.

5. **Promote activities of the Data Science Hub and campus data science ecosystem**
“Data Science Updates,” the Data Science Hub’s newsletter reaches almost 1400 subscribers with data science opportunities. The Facilitators present at conferences on campus and nationally to share the Data Science Hub’s work.