Data @ Whitman

PIG Summer 2025

Participants: Sharon Alker, Amy Blau, Will Boyles, Neal Christopherson, Arielle Cooley, Nancy Day, Michelle Janning, Rosie Mueller, Marina Ptukhina, Kevin Rineer, Dalia Biswas, Albert Schueller, Peter Shultz, Wiz Sugiarto, Chris Wakefield, Jordan Wirfs-Brock.

Introduction

Data is everywhere and integral to our lives. Increasingly, our students, our research, our alumni, and our institution rely on data for education, scholarship, careers, and the development of the College.

The time has come for us to organize and support the varied understandings, applications, and framings of data at the College. A group of faculty and staff first discussed this need during the 2024-25 academic year. Recognizing that clearly defining such an initiative would take more than a few meetings, the group applied for and received PIG funding to hold a series of workshops over the summer.

The goals of the workshop were:

- to enable the project team to decide what form a center for data should take at Whitman College and;
- to design a proposal that leaders in this initiative can begin introducing to a wider range of faculty, students, and staff to gauge other interests.

Assuming that there is adequate interest, beyond the PIG, we want to determine the shape of this initiative. Is it a center? Does it have staffing? Is it curricular, extracurricular, or both? Can it be started with existing resources? Can it be a truly interdisciplinary space?

In this report, based on the work of the faculty and staff that participated in the PIG, we offer a first draft to the community for what data at Whitman might look like.

A Center for Data

Here, we outline in broad terms what such an interdisciplinary, data center initiative might look like if we had funds to support it. Each of these emerged as frequent themes in our summer workshop.

Physical Space

The activities of the center should be focused on a specific, accessible, and visible space. Akin to the COWS or the STEM Hub, the center should have a dedicated space on campus.

Staffing

Director

The center should have a director. This person could be a faculty member with course releases or a full-time administrator. The director should have some administrative support in the form of an assistant.

Board

The director should have a small board of faculty/staff/students to help with policy and management of the center.

Steering Committee

The center should have a larger steering committee comprising faculty, staff, students, community, and alumni. The board and director should report to the steering committee a few times a year. The steering committee will foster relationships with stakeholders and help with fundraising.

Student Employees

The center should employ students for work, including, but not limited to, events, data and statistical consulting, data workshops, and tutoring.

Activities

- Resource for faculty, staff, and students working with data

- Ensuring these resources are calibrated to all levels of expertise, be it those who teach data science, or those who wish to understand data science
- Student-led consulting center supervised by a faculty member or senior student, similar to the Data Science Student Fellows Program at Bucknell
- Community engagement, manage consulting, finding solutions to real data questions
- Invite speakers, run a regular data seminar
- Manage experiential learning, data projects, sharing of experiences, possibly curricular in nature.
- Provide public-facing accounts of community activities, perhaps develop a regular Walla Walla data factbook, or contribute to the College's own factbook.

Ongoing Activities

What follows is somewhat supplemental to the first few pages of the report. This summary shows what the participants of the workshop are currently doing ("now") in the area of data and what they can do in the next academic year ("soon") in a resource-neutral way. This shows the breadth of data-centric activities on campus and highlights the need for investment in campus-wide coordination around this theme.

What We're Doing Now

Teaching

This section highlights the integration of data analysis, visualization, and various software tools into the curriculum across multiple departments.

- **Software Integration**: Courses are incorporating a wide range of data-focused software, including **R** and **R Markdown**, **Excel**, **Stata**, **Python**, **Matlab**, and qualitative tools like **NVivo** and **Dedoose**.
- **Skill Development**: There's a strong focus on building foundational data skills. This includes data analysis and visualization in chemistry and biology labs, survey design and analysis and presentation in sociology, and creating workshops for psychology students to practice data visualization in Excel.

- **Course-Based Projects**: Many classes feature hands-on projects, such as senior theses that are entirely data-based, an "Intro to Data Science" course where students analyze their personal data from services like Spotify and Google, and statistics projects where students create or find their own datasets.
- **New Courses & Curricula**: New courses have been created, including "Data Visceralization" which explores multi-sensory ways to communicate data through animation and sound, and a special topics course on databases taught in SQL.

Research

Faculty and students are engaged in a variety of data-driven research activities, supported by campus infrastructure and interdisciplinary collaboration.

- **Infrastructure & Support**: The college has a **High Performance Computing Linux Cluster** for intensive data processing. Additionally, faculty receive support for data management, storage, and recovery from the library and WCTS.
- **Interdisciplinary Partnerships**: Collaborations are key, such as a partnership between Biology and Computer Science to build image analysis tools for studying flower patterns. We would welcome partnerships with departments that may critique data, such as between Computer Science and a Philosophy course on the Ethics of Data.
- **Community-Engaged Research**: Faculty conduct program evaluations and needs assessments for local organizations in Walla Walla and the surrounding region, covering sectors like health care, housing, local government, education, and the arts. Faculty and student teams also work with local community partners on projects requiring digital data applications or access to data that help local organizations enhance their work.
- **Innovative Methods**: Research includes developing tools for "sonifications" (representing data with sound) and "physicalizations" (data sculptures), using spatial analysis in R, and writing a book on just and inclusive data-driven design.

Community

The efforts in this section focus on applying data and technology skills to projects that engage with the local Walla Walla community and beyond.

- **Local Partnerships**: There are active collaborations with community organizations, including a data project for the **Walla Walla Early Learning Coalition**, a design project with **Lincoln High School** students, and helping to organize agricultural data for the Walla Walla basin.
- **Educational Outreach**: Efforts include teaching an English course at the Washington State Penitentiary, building math activities for science outreach in local elementary schools, and a program to train students in Python with plans to expand to local schools and Bangladesh.
- **Public Engagement**: Data is being shared with the public in creative ways, such as creating an animated infographic on wine styles for a local winemaker and performing in data sonification concerts.
- **Consulting**: The College provides consulting services, such as Institutional Research, offering assessment help to other institutions and faculty consulting on "people research" with local groups.

Support Efforts

This section details the institutional framework and services that support data-related work for faculty, staff, and students.

- **Workshops & Training**: Institutional Research provides **Excel skills workshops** for staff, and faculty have offered workshops on managing research data.
- **Library & Data Services**: The library supports students and faculty by helping them find and cite data for theses and other research projects, supporting the use of ICPSR datasets, and offering occasional workshops on data management and preservation. This year, they will organize events for Love Data Week which is February 9-13.
- **Technology & Infrastructure**: WCTS (Whitman College Technology Services) manages licensing for software like **Matlab**, and there is ongoing work to improve data storage solutions on campus.
- **Collaborative Spaces**: The **STEM Hub** is highlighted as a welcoming and supportive space that fosters productive interactions between students and faculty.

Curricular Development

This section outlines the strategic, high-level efforts to create new academic programs and revise existing curricula to be more data-centric.

- **New Programs**: Major developments include the establishment of a **Data Science Minor**, the creation of a Human-Centered Design (HCD) concentration, and ongoing discussions to create a **Public Health** program, which will be highly interdisciplinary and data-focused. A combined Chem-CS major is also in development.
- **Curriculum Overhauls**: Several departments are formally revising their curricula (e.g. focusing on the use of R instead of relying on a lot of different technical tools). The psychology department is discussing its major with a focus on improving students' quantitative and statistical literacy. The sociology department undertook a PIG (Professional Interest Group) to better align quantitative skills across its courses, which included creating new Excel training videos.
- **New Courses**: New, foundational courses have been created, such as "Intro to Data Science" and "Applied Machine Learning".

Communications

This area focuses on sharing data-related work and knowledge with various audiences.

- **Speaker Series**: The **Math Science Foundry Series** provides a platform for researchers to present their work to students.
- **Publications & Workshops**: Faculty are publishing their work, including a book chapter on a community-based design lab project, and giving external workshops on using data for organizational decision-making.
- **Institutional Reporting**: The **Whitman College Factbook**, published by Institutional Research, serves as a key communication tool for sharing institutional data.

What We Can Do Soon

Summary of what we can do in the near term.

Community Engagement and Partnerships

- There is a suggestion to pursue community partnerships with groups like local Tribes or the wine industry to make data sets available to the public, possibly with a grant. This work can be in cooperation with CELRI, or via other avenues.

- It was noted that nonprofits would be an ideal group to partner with. A proposal was made to gather local nonprofits to gauge their interest in a data center and what programming would be useful to them.
- One idea is to use independent studies for departmental outreach to the wider Walla Walla community. A speaker series could also be retooled to focus more on community data.
- An offer was made to give a talk on designing data projects with community partners and building those relationships.

Curriculum and Course Development

- A proposal was made for a new project-based course, "Data Across Disciplines," which could be co-taught by faculty from different fields and would have no prerequisites to encourage non-STEM students to enroll.
- It was suggested that a course like this could lead to a "+Data Inquiry" addition to majors in the future and might also relieve some pressure on the Intro to CS course. There are also plans to develop an Intro to Python course that would be project-based and serve as an alternative to Intro to CS for students in quantitative fields.
- There is a plan to incorporate more data and foundational skills like file management and Excel into the Principles of Macroeconomics course.
- Potential new statistics courses are being developed, including "Correlated Data" and "Statistical Learning." There is also a plan to work with library staff to curate datasets from the archives for students and faculty to use in classes.
- It was suggested to make linear algebra a course with no prerequisites.
- One of the Professional Interest Groups is working on helping students gain statistical and technical skills for research methods.
- Cross disciplinary independent studies could be taught that have faculty supervisors with different areas of expertise - one being a content expert in a humanities field, for example, while the other may have expertise in data-based tools.

Workshops and Student Support

- An offer was made to lead a workshop for faculty and students on qualitative data analysis using tools like pen-and-paper and Dedoose. Dedoose is a cloud-based app that was developed by UCLA researchers. It allows you to analyze text, video, audio, and PDFs to uncover hidden patterns in your data.
- Proposals were made for a basic Excel workshop for staff and a "Storytelling with Data" workshop.

- There are plans to organize short workshops for students on skills needed for data science, such as GitHub, Overleaf, R Projects, and Jupyter Notebooks.
- It was suggested that the Econ department's tutors for R could be advertised more broadly to assist students across campus.
- A pilot program was proposed where students who have already taken and tutored for a lab course could hold STEM office hours to consult with other students, creating a model for a future data center.
- It was suggested to offer tech seminars for new students, possibly during orientation, to improve the new-student experience.

Infrastructure and Resources

- There is a plan to identify recommended data storage providers for various research scenarios.
- Work is being done to integrate a web portal called "Open OnDemand" developed by the Ohio Supercomputer Center for the High Performance Computing Cluster to provide students with familiar interfaces like RStudio and Jupyter. There are also plans to set up a GitHub organization for students to create portfolios of their work.
- A general suggestion was made to identify a physical space on or off-campus that could serve as a consulting center or a space for "data analysis pop-ups."

Institutional and Faculty Collaboration

- There are plans to reach out to CELRI and humanities faculty to broaden the conversation.
- There is a plan to showcase faculty use of data for the Whitman community in a series of lightning talks; there will be outreach across the faculty to find participants
- The possibility of a reading group, which could include students, will be explored.
- It was suggested to collaborate on creating a two-page strategic document to highlight the data-related work already being done, which could be shared with donors, trustees, and students.
- Funding sources for these initiatives will be explored.
- The ad hoc Classroom Committee is exploring learning spaces generally, with an eye toward initiatives such as this one that require spaces on campus for their work. Faculty involved in this PIG can offer ideas for what kinds of spaces may best meet the needs of this kind of initiative.

- A general agreement was made to continue the conversation with regular check-ins for both small and large groups throughout the semester.