

Public-land disputes in the American West touch on everything from Indigenous rights to conservation to the placement of wind farms, and yet, much of the existing academic work still leans heavily on small case studies or anecdotal observations. Over the past academic year, I've been fortunate to work on a project that tries to gather more information on public land disputes in the United States. Under the guidance of Professor Cheryl Schonhardt-Bailey, I helped build a large, systematically coded database of public-land legal disputes decided by the U.S. Ninth and Tenth Circuit Courts of Appeals. The goal was to bring these scattered cases together into a single, structured resource.

My role in the project unfolded in three main stages. The first was manual coding: I read and coded 600 individual court decisions, logging around 20 data points per case, including things like the types of parties involved, which circuit the case was heard in, the outcome, whether there were counterclaims, and specific Westlaw headnotes. This became the "gold-standard" dataset that we later used to train an AI model.

The second stage was scaling up. We initially experimented with Google's Gemini API to automate the coding process, but eventually switched to OpenAI's API for better reliability. I wrote Python code in a shared Google Colab notebook to run the model, allowing our team to keep everything transparent and version-controlled. By the end of this stage, we had automatically processed over 6,000 additional cases.

Finally, I moved into data cleaning and validation, importing everything into R, standardising fields, removing duplicates, and cross-checking the AI outputs against the hand-coded entries. After several rounds of cleaning and quality control, we ended up with a final dataset of nearly 5,000 usable cases.

By providing structured data on thousands of court decisions, the database offers a solid foundation for anyone interested in public-land governance. While we had specific research goals in mind, I think the real value of this work is that it gives future researchers a much broader and more reliable base to build from.

Personally, this project was incredibly rewarding as I improved not only my hard skills but also soft skills in research. Coming in, I had some intermediate experience with Python. I was able to improve my coding throughout the project, and by the end of this programme, I was confidently writing code to interact with APIs and process data at scale. I also found myself diving deep into a legal field I knew very little about, public-lands jurisprudence. The regular check-ins with Professor Schonhardt-Bailey were especially valuable. They taught me that research isn't always a straight line, and that being flexible and responsive is just as important as being thorough.

This project allowed me to summarise complex legal issues into structured data, contribute to original research, and gain firsthand insight into how meaningful academic work is built from the ground up. I'm incredibly grateful to the Phelan US Centre, Professor Schonhardt-Bailey, and the generous donor who made this opportunity possible. It's been an experience that's not only shaped my academic interests, but also given me a clearer sense of where I might want to go next.