**Reflection Open Science**

Emily Rivera Mondragon
December 2025

## Introduction

For my final project, I chose to publish a small open-access dataset related to my dissertation research on the factors influencing the adoption of the Global Covenant of Mayors for Climate and Energy (GCoM) among Mexican municipalities. I selected this option because it directly supports the long-term goals of my research while allowing me to practice meaningful open science principles.

## What I Did

The first step in my project was defining a dataset that could be safely shared without violating confidentiality or restrictions. My full research dataset includes public information from National Institute of Statistics and Geography (INEGI) 2010 and National Council for the Evaluation of Social Development Policy (CONEVAL) 2010, but it also incorporates sensitive government data from Mexican Petroleum (PEMEX) that cannot be made public. To comply with ethical and legal considerations, I created a small representative dataset that includes only publicly accessible variables: population, poverty percentage, higher-education rates, GCoM membership status, and an indicator of civic climate mobilization. These variables are sufficient to illustrate the key socioeconomic and civic dynamics relevant to my research questions without revealing protected or restricted information.

Once the variable selection was finalized, I organized the data into a clean, easy-to-read CSV file named gcom-mexico-sample-dataset.csv. I uploaded this file to a new public repository on GitHub. The repository also includes a detailed README file explaining the purpose of the dataset, definitions of each variable, data sources, license information (CC-BY 4.0), and recommended citation. Additionally, I created a metadata.json file following basic metadata standards to ensure that anyone downloading the dataset understands its context, content, authorship, and terms of use.

This process required me to think about transparency in ways I had not previously considered. Instead of focusing only on the data itself, I had to document the steps, decisions, and limitations behind it. By doing so, I created a small yet complete open-data product that others could reuse, critique, or expand upon.

**Why I Chose This Option**

I selected the dataset publication option because it aligned most closely with my dissertation needs and because I will continue working with data for the next several years. Publishing a small dataset was both practical and impactful: it added real value to my research workflow without creating extra work that does not benefit my long-term goals.

I also wanted to gain experience with GitHub, since it is widely used in environmental research, data science, and climate-policy analysis. Before this course, I was not comfortable using GitHub, and I assumed it required advanced programming skills. Completing this project helped me realize that GitHub can be used effectively even by beginners, and that uploading and documenting datasets can be done entirely through its web interface. This gave me more confidence to use GitHub as an open science tool in future projects.

In addition, this option allowed me to reflect on the importance of making at least part of my research more accessible. Much of the literature on climate governance emphasizes transparency and public participation, and by sharing data openly, I am modeling the same principles in my academic work. Publishing even small datasets can help other researchers, policymakers, or students understand barriers and drivers of climate program adoption in Mexico.

**How This Supports Open Science**

This project advances open science in several key ways:

- **Accessibility**: The dataset is openly available on GitHub, reducing barriers for students and researchers with limited resources.
- **Reproducibility**: Sharing a well-defined subset allows others to examine the structure of the data, understand my variables, and replicate parts of my methodological approach. The metadata and README ensure clarity.
- **Preservation:** GitHub provides a stable platform where future versions of the work can be shared, corrected, or expanded, ensuring sustained accessibility.
- **Ethical data sharing:** I balanced openness with ethical considerations, ensuring that transparency does not compromise privacy or confidentiality.
- **FAIR & Licensing:** The dataset follows FAIR principles (Findable, Accessible, Interoperable, Reusable) and uses a CC-BY 4.0 license to encourage proper citation.

**Conclusion**

Overall, publishing a small open-access dataset was a meaningful and practical way to apply open science principles to my research. It improved the clarity, accessibility, and structure of my work while helping me develop new technical skills. Most importantly, it reinforced that open science is not only about sharing data, it is about promoting transparency, trust, and collaboration within the research community. I plan to continue incorporating these practices throughout my PhD and future academic career.

Link to repository: https://github.com/EmilyRivera7/gcom-mexico-sample-dataset