

Fondren Fellows

Project Title

Open Data, Open Knowledge: Rice University Python-Based Solutions for Enhancing Access and Literacy in AI Era

Brief Summary

This project makes data more accessible by creating a Python tool which can automatically collect, update, and sift through datasets which are normally stored in very unstructured ways on the internet. Additionally, we will provide code and documentation on how to use our tool and, further still, how to derive some conclusions and visualizations from the data using fundamental data science techniques. These tools will have written documentation and a comprehensive training materials package to demonstrate to the Rice community how to use them. By improving access to difficult-to-use data, we enhance its usefulness to the Rice research community.

Project Description:

Artificial intelligence (AI) and machine learning (statistical models) have become very popular and powerful tools, and have been widely used in many domains, such as social science, education, politics, etc. Consequently, more and more students, faculty, and staff at Rice have started studying and applying AI in their studies, research, and work.

However, data access and compilation has become a significant barrier against people using AI for two reasons. First, most public data websites lack a friendly interface for people to easily access and compile large volume data required for many AI techniques. Second, it is challenging for people to receive sufficient software and coding training for accessing and compiling large volume public data from conventional curricula in many departments.

To remove these barriers, Rice Kelley Center collaborating with Rice Data To Knowledge (D2K) Lab propose this project for achieving the following three objectives:

- **Data access software** makes access to public datasets easier for the user.
- **Data visualisation software** facilitates user visualization and understanding.
- **Training** allows students, faculty, and the general population to use these tools.

The impact of this project can be summarized in three levels:

- **Individual level** - this project will support students, faculty, and staff in all kinds of domains by making data more accessible.
- **Department level** - this project will enhance Kelley Center providing service for accessing and using public government data, and promote D2K Lab conducting data science education in different departments.
- **Rice community and beyond** - this project will provide these tools publicly, making them useful to anyone in the world who has a need.

Our team will collaborate closely to implement this project. Specifically, Dr. Luzi and Dr. Lan are two assistant teaching professors from D2K, and they will supervise Fondren fellows designing the software for public data access and visualization. Ms. Xiong, the government

information coordinator at the Kelley Center, will oversee product testing, training materials package, and workshop series (if time permits).

We will initially design data access and visualization code based on the Texas Education Agency (TEA) website, which is a representative website without friendly data access. TEA data is widely used by many students, faculty, and staff at Rice. As the project continues, we plan to

include more public data sources based on project progress and the feedback from the Rice community.

The final project deliverables include software packages and training materials, and have the following properties:

1. The software and training materials will be designed based on the Python language and targeted for users with basic coding background.
2. The training materials include documents, slides, demos, and workshops or other alternative forms of instruction covering all the details for users applying the designed software packages for data access and visualization.
3. All software packages will be accessed via Fondren library and a public Github repository.
4. The software packages include two versions: standardized Python packages for everyone and customized public interface for specific Rice departments.
 - a. The standardized Python packages enable common users to
 - i. check and download data with the latest version from most government websites by configuring some parameters, e.g., target dataset website, time range, locations, variables of interest, etc.
 - ii. compile data in a structured format for existing machine learning modules to process, and visualize the main information of downloaded data by simple configurations.
 - b. The customized public interface allows specific Rice departments to download and visualize data by simple operations in the customized software panel.

Key Tasks

- Developing python code for crawling data from public website
- Developing python code for visualization
- Preparing training material package for the Rice community and general public
- Maintaining project notes and meeting minutes.

We expect to hire two Fondren Fellows. One Fellow will focus on designing public data access and visualization packages, while the second will concentrate on testing, providing feedback, and developing user-friendly training material packages. The Fellows' work will be supervised through monthly mentor meetings, supplemented by weekly one-on-one meetings between each mentor and mentee, along with additional meetings involving external collaborators.

Beginning of the Fall 2025 Semester:

- Launch recruitment and training for Fondren Fellows.
- The entire team will meet to brainstorm and establish specific goals, quality testing criteria, work distribution, team workflow, action items, timeline, and milestones.

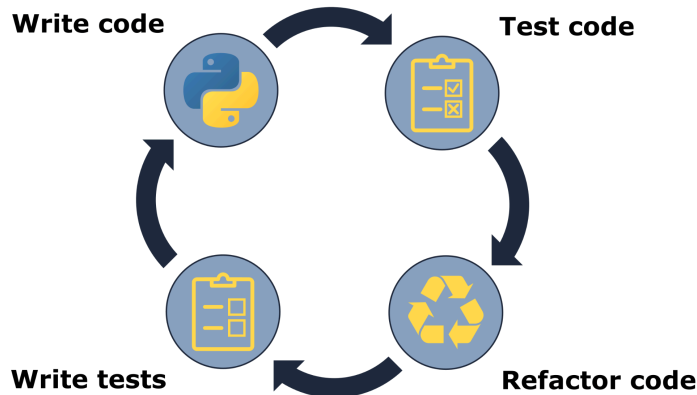
Fall 2025 – Spring 2026:

- **Design Team:** The Fellow, mentored by Dr. Luzi and Dr. Lan, will focus on developing and refining public data access and visualization Python software packages. Their work will involve an iterative process of development, testing, and documentation to

ensure the final product is fully functional and accessible to the public.

- **Testing & Training Fellow:** The second Fellow will conduct ongoing testing and provide feedback to ensure the product is easy to use, especially for those users who have basic programming knowledge and skills. The Fellow will also create a comprehensive user

friendly training material package to facilitate user adoption.



Under the mentorship of Ms. Xiong, a Fellow will oversee the testing process for each component developed by the design team, including technical documentation, and provide feedback for improvement. The Fellow will:

- Plan and design a user-friendly training materials package, incorporating documentation and instructional resources that do not require workshop attendance and require basic programming knowledge.
- Seek feedback from the design team and refine the materials accordingly.
- Recruit volunteers from targeted user groups—including students or social science researchers (e.g. Rice HERC researchers) with minimal data programming experience, as well as librarians responsible for data training, reference services, and research consultation—to test our product and training materials to gather additional feedback.
- Analyze and integrate feedback to revise and finalize the training materials.
- Optional: Organize and conduct a series of data training workshops if time permits

End of the Spring 2026 Semester:

By the end of the 2026 spring semester, the Fellows will launch the product and training material package for public use. Toward the semester's conclusion, both Fellows will also reflect on their work, write reports, and collaborate on a short project presentation for the Fondren Fellows showcase meeting.

What qualifications would you expect from students working on this project?

Basic requirements:

- Self-motivated and hard-working
- Teamwork spirit
- Python programming
- Teaching and/or research experience

Preferred skills:

- Web crawling with Python
- Javascript for web deployment
- Data visualization based on Tableau, Power BI, or other commonly used softwares
- Special knowledge about issues involved with using government datasets for social science research
- User experience assessment
- Special knowledge, skill and experience with curriculum and instruction design, and writing tutorials
- Be able to use tools for graphic design, creating video clips

What would students learn through their participation in this project?

Students will mainly learn the following skills from this project:

- Python coding skills
- Data visualization
- Technical presentation and writing
- The practical skills for teamwork and collaboration
- Archiving data science project and sharing it through open source
- Developing data training material package
- Possibly organize and conduct a series of data training workshops if time permits