

PUBLIC DATA RESEARCH BARRIERS

Two barriers keep researchers from using public datasets:

- Government websites don't make it easy to find or download files
- Manually hunting for files across hundreds of pages is slow, technical, and frustrating

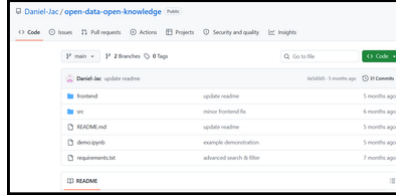
OUR PILOT TOOL for removing the barriers

Backend: Shareable and Editable Codes



WEB DRAWLER IN 3 SIMPLE STEPS

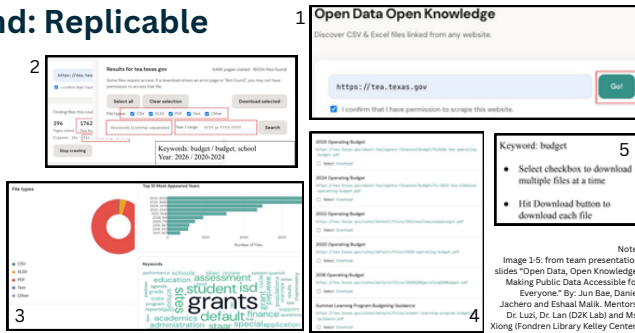
- 1. PASTE A URL**
 - Enter any public website and confirm consent
- 2. CRAWL & DISCOVER**
 - Crawls a website and surfaces every data file it finds — CSV, XLSX, JSON, PDF.
- 3. FILTER & DOWNLOAD**
 - You can set parameters to search, filter, and download what you need.



Front End: Replicable

KEY FEATURES & RESULTS:

- Search and filter results
- by file type, keyword, or year
- Track website search/crawl progress in real time and stop time
- Explore findings through visualizations



TRAINING GUIDE: LEARN. RACTICE. SUCCEED.

Overview

- Developed as a [LibGuide](#)
- Designed to make a technical tool accessible to a wider audience
- Provided a structured learning experience

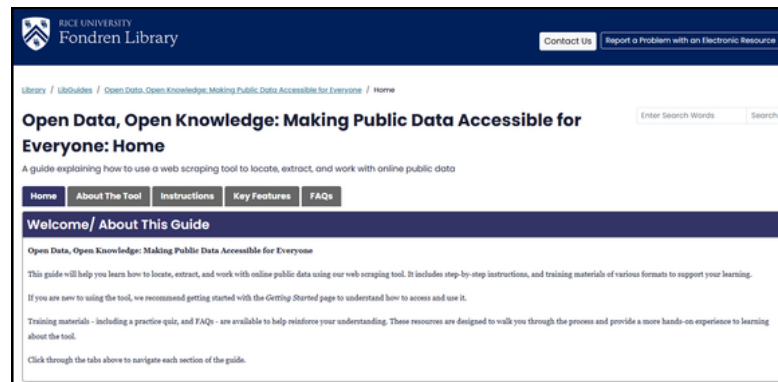
Multiple learning formats for different styles and needs

- Short video tutorials for key tasks
- Step-by-step written instructions
- FAQ page for common questions in clear, easy-to-understand language

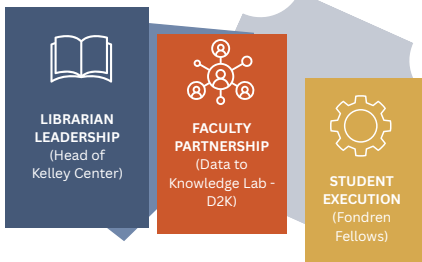


Who it's for

- students, researchers, teachers, school administrators, and IT professionals, and anyone who is interested in public data



HOW DID WE ACHIEVE IT



A Triad of Expertise: Project is structured around three core pillars under the Fondren Follows Model, utilizing student fellows to drive innovation under librarian and professors' guidance.

Strategic Mentorship Roles: Librarian provided domain expertise in data resources and data users: Faculty provided technical oversight for student-led Technical and Training teams.

BENEFITS FOR STAKEHOLDERS

FOR LIBRARIANS:

- Catalyst to transform library from passive storage place to data research partner and experiential learning place

FOR FACULTY:

Applied Research Integration

- Provides a platform for teaching professors in Data Science to lead student researchers apply theoretical knowledge to overcome real-world accessibility challenges

FOR STUDENTS:

- Tech & Civic Education Leadership
- Experiential learning
- hands-on experience in Python development, web scraping, and public education platform like LibGuides, friendly documentation and training materials for users with different tech skill levels

PHASE 1: TEAM FORMATION & FUNDING

September: Secured fellowship funding, assembled the Mentorship Team, and recruited undergraduate student research fellows.

Technical Team, mentored by D2K faculty, built core Python scraping solutions to handle complex government data sources and provided clear documentation for programmers. Training team mentored by librarian tested and provided feedback.

PHASE 2: TECHNICAL DEVELOPMENT

Training Team developed the "Open Data, Open Knowledge" LibGuide, ensuring technical tools are accompanied by clear instructions for everyone, especially non-programmers.

PHASE 3: EDUCATIONAL DESIGN

Finalized the project (April/May 2026), launched the tool and published the LibGuide for public use and published sharable codes on GitHub for public goods

PHASE 4: LAUNCH & SHARE

DATA LITERACY AT NEXT LEVEL

- Focusing on data literacy to ensure government and non government public data are not just "open" but accessible for use in data research and analysis.
- Librarian as catalyst. DS faculty mentor CS students apply tech skills to solve real problems and developing training materials for public goods.

FUNDING & MENTORSHIP TEAM

Funded by 2025/2026 Rice University [Fondren Fellows Program](#)
Mentors: Ms. Anna Xiong, Dr. Lorenzo Luzi, & Dr. Xinjie Lan
Project contact: Anna Xiong, jax2@rice.edu

STUDENT EXECUTION TEAMS

Tech Team: Jun Bae, Daniel Jachero
Training Team: Amy Xing (Fall Semester), Eshaal Malik (Spring Semester)