



POLICY DESIGN & EVALUATION LAB

Research Transparency
Initiative

Principal Investigator: Craig McIntosh

UC San Diego
School of Global Policy and Strategy



PDEL Data Publication Services - Motivation

1. Researchers often don't have the time or resources needed to clean, annotate and publish the data files at the conclusion of their research.
2. Data publication is a crucial part of conducting transparent and reproducible social science.
 - Bolsters quality of research by permitting replication studies.
 - Expands investments made in generating new data by allowing others to extend the data in new dimensions.
3. There is great value to PIs in having someone external to the research team replicate data analyses and cross-check with results in the paper.



PDEL Data Publication Services – Aims

1. Take datasets that are not currently in the public domain; anonymize and annotate them; work with UCSD Library to enter them into the Digital Collection.
2. Form a campus cluster around data science and replicability.

Progress on Data Publication

Completed collections

Principal Investigator(s)	Title	Full Citation
Adida, Claire L; Ferree, Karen E; Posner, Daniel N; Lea Robinson, Amanda	Data from: Who's Asking? Interviewer Coethnicity Effects in African Survey Data	Adida, Claire L; Ferree, Karen E; Posner, Daniel N; Lea Robinson, Amanda (2016): Data from: Who's Asking? Interviewer Coethnicity Effects in African Survey Data. UC San Diego Library Digital Collections. http://doi.org/10.6075/J06H4FBV

Pending Collections	
Principal Investigator(s)	Title
Craig McIntosh	Data from: Ethiopian Project on Interlinking Insurance with Credit for Agriculture (EPIICA)
Gollin, Douglas, David Lagakos, and Michael E. Waugh	Data from: The agricultural productivity gap
Craig McIntosh	Data from: Schooling, Income, and HIV Risk in Malawi (SIHR)
Ferree, Karen E., Gibson, Clark, Jung, Danielle F., Long, James D., McIntosh, Craig	Data from: How Technology Shapes the Crowd: Participation in the 2014 South African Election
Hanson, Gordon and McIntosh, Craig	Data from: Is the Mediterranean the New Rio Grand? U.S. and E.U. Immigration Pressures in the Long Run
In Progress	
Principal Investigator(s)	Title
Adida, Claire, Jessica Gottlieb, Eric Kramon and Gwyneth McClendon	Data from: When Good News is Bad...and When it Isn't: Voter Coordination, Preferences, and Electoral Behavior
Jesse Driscoll	Multiple datasets
Thad Kousser	Multiple datasets
Scott Desposato	One dataset

Total UCSD Faculty members supported: 9, from three departments



Forming a cluster around research transparency at UCSD

Training in best practices in reproducible social science

- RAs attended the Berkeley Initiative for Transparency in the Social Sciences (BITSS) annual meeting.
- Co-sponsored a hands-on workshop by Garret Christensen (BITSS project scientist) in implementing a reproducible workflow.

Prepared a major NSF grant to support an innovative graduate education program.

- Program contains a significant element of training in data science and reproducibility.
- Involves 8 faculty members across 7 departments and is open to all UCSD graduate students.



UCSD in the broader research transparency movement

- Created GitHub wiki pages to document PDEL data publication processes.
 - Shared with colleagues at MIT and the Center for Effective Global Action who are considering offering the same services to their PIs
- Writing a guest blogpost for BITSS on the UCSD process and problems encountered.
- Participating in a UC Multicampus Research Initiative proposal to establish the California Institute for Reproducible and Transparent Science (CIRTS)

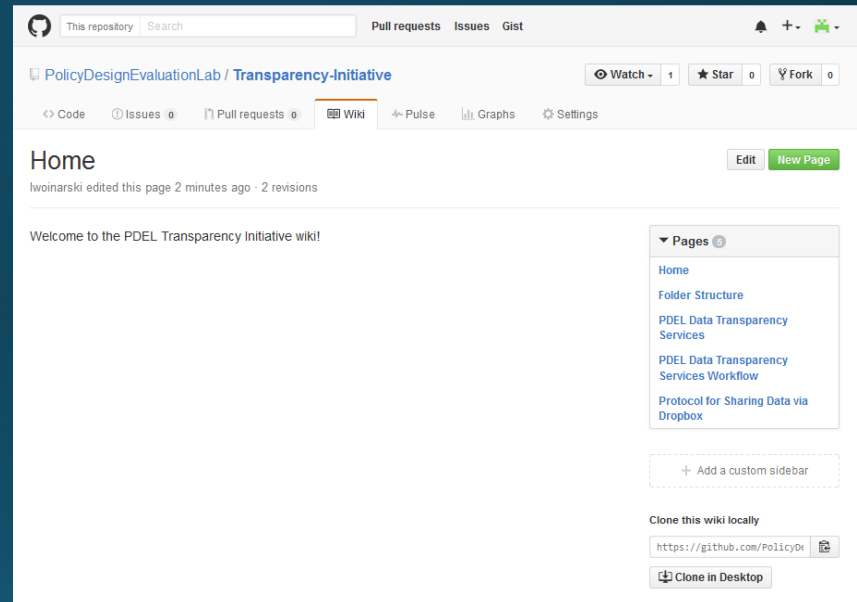
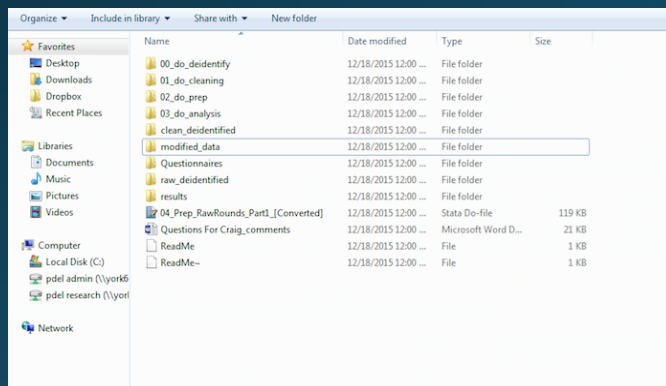
PDEL Transparency Initiative GitHub Wiki pages

1. PDEL Data Transparency Initiative – Data Publication Services

Description of the motivation, aims, services, and requirements.

2. Folder Structure Best Practices

Suggested structure for organizing data and code related to a research project in a way that others can more easily understand and replicate your project.



3. Workflow for prepping data for publication

Detailed step-by-step description of the process to streamline code, anonymize data, organize files, and replicate analyses.

4. Protocol for Sharing Data via Dropbox

<https://github.com/PolicyDesignEvaluationLab/Transparency-Initiative/wiki>