

Preparing to Share Social Science Data: An Open Source, DDI-based Curation System

Limor Peer

Yale University Institution for
Social and Policy Studies (ISPS)

Stephanie Wykstra

Innovations for Poverty Action (IPA)

Jeremy Iverson & Dan Smith

Colectica

Ann Green

Digital Lifecycle Research & Consulting

Corresponding author email: limor.peer@yale.edu

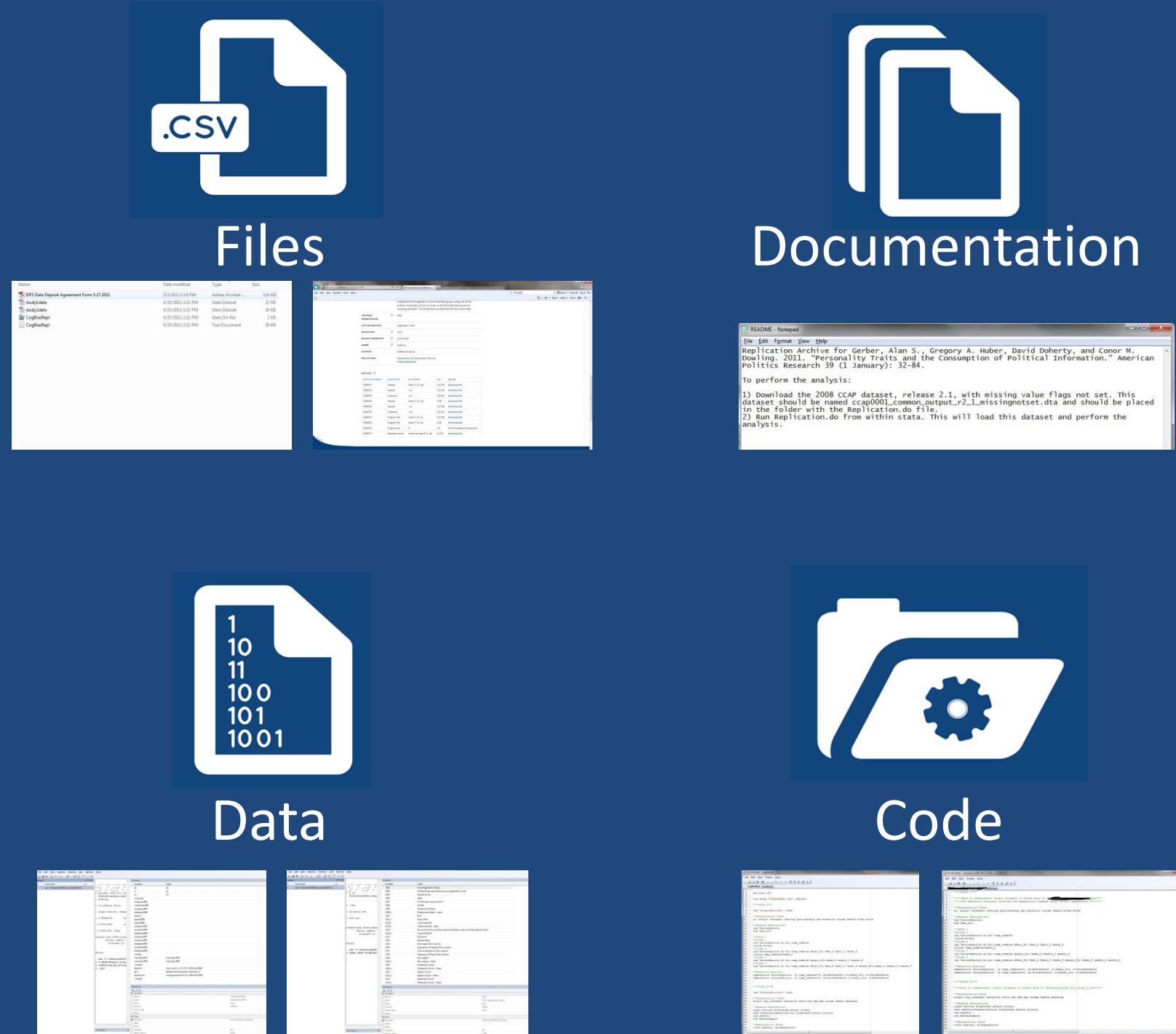
The value of shared data is linked to its usability; sharing data that is not usable in the long term is ineffective. But how to support the preparation of research materials for replication, reproducibility, repurposing, and reuse? This new data curation software includes curation steps designed to improve research materials and to enable users to derive greater value from the data by performing tasks such as checking variable-level and study-level metadata, replicating code to reproduce published results, ensuring that PII is removed, and creating preservation formats. The tool is based upon the best practices of data archives and fits into repository and research workflows.

Guiding principles

“Ensure that the information to be preserved is Independently Understandable to the Designated Community. In particular, the Designated Community should be able to understand the information without needing special resources such as the assistance of the experts who produced the information.” (OAIS, 2012)

“Intelligent openness: data must be accessible and readily located; they must be intelligible to those who wish to scrutinise them; data must be assessable so that judgments can be made about their reliability and the competence of those who created them; and they must be usable by others.” (Royal Society, 2012)

Data Quality Review



See: Peer, Green, and Stephenson. 2014. Committing to Data Quality Review. International Journal of Digital Curation, 19(1): 263-291. doi:10.2218/ijdc.v9i1.317

About the project

The Institution for Social and Policy Studies (ISPS) at Yale University and Innovations for Poverty Action (IPA) are partnering with Colectica to develop a software platform that structures the curation workflow.

Both organizations collect data from social science research that measures the impact of interventions—such as voter mobilization campaigns and microfinance programs—via randomized controlled trials in the real world.

The software leverages DDI Lifecycle and combines several off-the-shelf components with a new, open source, Web application.

ISPS and IPA requirements

- Curation workflow management (dashboard)
- Track changes to files (provenance)
- Integrate metadata production with data and code review and cleaning
- Preservation metadata and formats
- Secure storage and access
- Smooth transition to public dissemination of content
- Preference for open source solutions

