

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

Limor Peer, Yale University Institution for Social and Policy Studies
Ann Green, Digital Lifecycle Research & Consulting
Stephanie Wykstra, Innovations for Poverty Action
Jeremy Iverson, Colectica
Dan Smith, Colectica

Submitted to the 2015 e-science symposium, 9 April 2015

Abstract

Keywords: digital curation, data repository, research data, data sharing, reproducibility, social science research, open access

Objective: This poster will describe the development of a curatorial system to support a repository for research data from randomized controlled trials in the social sciences.

Description: 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, including checking data for confidentiality and completeness, creating preservation formats, and reviewing and verifying code. The software leverages DDI Lifecycle – the standard for data documentation – and will enable a seamless framework for collecting, processing, archiving, and publishing data. This data curation software system combines several off-the-shelf components with a new, open source, Web application that integrates the existing components to create a flexible data pipeline. The software will help automate parts of the data pipeline and will unify the workflow for staff, and potentially for researchers. Default components include Fedora Commons, Colectica Repository, and Drupal, but the software is developed so each of these can be swapped for alternatives.

Results: The software is designed to integrate into any repository workflow, and can also be incorporated earlier in the research workflow, ensuring eventual data and code deposits are of the highest quality.

Conclusions: This poster will describe the requirements for the new curatorial workflow tool, the components of the system, how tasks are launched and tracked, and the benefits of building an integrated curatorial system for data, documentation, and code.