

Data Under Constraint: Tools and Strategies for Facilitating Transparency

Draft Workshop Overview and Agenda

Wednesday, August 30, 2017 | Hilton San Francisco | San Francisco, CA

Organizers: Data-PASS Alliance (<http://www.data-pass.org/>)

OVERVIEW

Journal editors across the social sciences are designing and introducing new transparency guidelines requiring authors to share the data that underlie their publications and explicate how those data were generated and analyzed. A key concern for editors and authors is how to deal with data that are under constraint – e.g., data that were generated through interaction with human participants, are classified, are under copyright, or have some other proprietary content. Authors of course must comply with relevant agreements they signed, ethical commitments they made, and applicable laws. How can journal editors facilitate transparency under these conditions, and how much transparency can they encourage? The workshop will focus on these critical questions and challenges.

The workshop is the third in a series on "Developing and Implementing Data Policies: Conversations Between Journals and Data Repositories." The series is designed to promote discussion among social science journal editors, personnel from data repositories, data librarians, and other relevant constituencies about current approaches to data citation, management, and archiving. As with previous events in the series, the workshop is being organized and led by various members of the Data Preservation Alliance for the Social Sciences (Data-PASS, www.data-pass.org), a consortium of social science data repositories: the Institute for Quantitative Social Science (IQSS) at Harvard University; the Howard W. Odum Institute for Research in Social Science at the University of North Carolina-Chapel Hill; the Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan; the Qualitative Data Repository (QDR) at Syracuse University; and the Cornell Institute for Social and Economic Research (CISER) and the Roper Center for Public Opinion Research at Cornell University.

AGENDA

Lunch (1:00pm – 1:30pm)

Introductions (1:30pm – 2:00pm)

Introductions, workshop outcomes, initial questions, distribution of attendees list (for networking and continued discussion on data under constraint).

Dataverse, Journals, and Sensitive Data (2:00 – 2:30pm)

Gustavo Durand, Technical Lead / Architect, Dataverse, Institute for Quantitative Social Science, Harvard University

Dataverse is an open source platform to build data repositories for publishing and citing research data. It facilitates making data available to others, and allows you to replicate others' work more easily. After first introducing Dataverse, this session will present current features that were built for and are used by journal editors, such as the Review workflow and Private URL sharing and delve into a deeper examination of the permission management system. Finally, this session will discuss the upcoming vision of Dataverse 5, which brings support for higher levels of sensitive data, including file-level security and access requirements and the integrations with DataTags, a system for classifying datasets according to their level of sensitivity, and PSI, a differential privacy tool.

Adapting Data Verification Workflows to Accommodate Restricted Replication Data (2:30 – 3:00pm)

Thu-Mai Christian, Assistant Director for Archives, Odum Institute for Research in Social Science, University of North Carolina, Chapel Hill

This session will describe the implementation of the *American Journal of Political Science* [Replication and Verification Policy](#), which requires authors to submit replication materials for independent verification to ensure that figures and tables presented in articles can be reproduced using author-submitted data and code. The session will provide details on the integration of manuscript submission and data verification workflows, with special attention given to special cases in which replication data are proprietary or confidential. These examples will demonstrate ways in which such obstacles can be overcome in order to fully support research transparency.

Coffee Break 3:00 – 3:15pm

ICPSR's Restricted-use Data Management and Virtual Data Enclave (3:15 – 3:45)

Justin Noble, Acquisitions Manager, Inter-university Consortium for Political and Social Research (ICPSR)

This session will highlight restricted-use data management at ICPSR, including ICPSR's Virtual Data Enclave (VDE). The ICPSR VDE allows secure access to sensitive or confidential data through a virtual private network connection to a portal on a desktop computer. This session will also introduce other services offered by ICPSR including delayed dissemination, the ability to share data with collaborators at various permission levels prior to publishing a project with ICPSR, and Institutional open ICPSR, a research data-sharing service developed to meet the needs of universities, journals, professional associations, research centers, and departments.

Using Roper Center Data to Satisfy Transparency Requirements (3:45 – 4:15pm)

Peter K. Enns, Executive Director, Roper Center and Associate Professor, Department of Government, Cornell University

Like those of many institutions that house quantitative data, the terms and conditions of the [Roper Center for Public Opinion Research](#) prohibit the dissemination of individual-level Roper Center data. Because Roper data are cited in almost 400 publications each year, this constraint could pose a challenge to journals and researchers who wish to make replication data available. However, even when data constraints exist, there are often ways to enhance data sharing and transparency. This presentation will show journal editors and researchers how the Roper Center can be used to easily satisfy data transparency requirements and how archiving replication data with the Roper Center increases the visibility of the research and replication data. We will also discuss how the Roper Center approach could serve as a model for other data that cannot be made public.

Break -- 4:15pm-4:30pm

Journal Editors Discussion Interface (JEDI) (4:30pm – 5:00pm)

Colin Elman, Director, Qualitative Data Repository (QDR)

The “Journal Editors’ Discussion Interface” ([JEDI](#)) facilitates interaction and exchange among the editors of social science journals. JEDI’s initial focus is on the parts of the editorial process that deal with data and their analysis (e.g., effective data management, data citation, and linking data and analysis to published conclusions), and making research more transparent. JEDI’s goal is to catalyze the online community necessary for the dialogue to become self-sustaining. We seek willing participants to engage actively in JEDI’s launch and to sustain its initial conversations.

Group Dinner for Workshop Participants (7:00pm)

Location TBA