A repository on a mission: A small research community gets serious about reproducibility

Limor Peer, ISPS at Yale University
Submitted to the 2012 e-science symposium, 4 April 2012

Abstract

Keywords: data archive, digital repository, research data, replication, reproducibility, social science research, open access

Objective: To describe the process and challenges of creating a replication data archive at the Institution for Social and Policy Studies (ISPS) at Yale University. The Archive provides open access to research data, links data to publications, and ultimately facilitates reproducibility.

Description: The ISPS Data Archive is a digital repository for research produced by scholars affiliated with ISPS, with special focus on experimental design and methods. The primary goal of the Archive is to be used for replicating research results, i.e. by using author-provided code and data. The Archive was launched in September 2010 as a pilot for Yale’s Office of Digital Assets and Infrastructure (ODAI) to find solutions relating to storage, persistent linking, long-term preservation, and integration with a developing institutional repository.

Results: Before data publication, Archive staff processes data and code files, including verifying replication, adding metadata, and converting to CSV and R. To date, the ISPS Data Archive has published over 750 files for about 45 studies.

Conclusions: The development and implementation of the ISPS Data Archive, though outside the library, raises issues familiar to librarians: the need for clear policies from the institution; the challenge of finding support for the provision of high quality services; the complexity of working in close partnership with IT; the need to keep up with fast-paced changes in technology and in user expectations; and the challenge of bringing about change in community norms and practices. Alongside these practical issues, fundamental questions arise about the appropriate role of the university vs. the disciplines when it comes to data archiving, especially in light of the need to comply with requirements from funders and journals.