



BERKELEY INITIATIVE FOR TRANSPARENCY
IN THE SOCIAL SCIENCES

Short Paper
Title

Christensen

Introduction

Replication

Project Protocol,
Reporting Standards

Workflow

Data Sharing

Data Sharing and Replication Transparency

Garret Christensen¹

¹UC Berkeley: Berkely Initiative for Transparency in the Social Sciences
Berkeley Institute for Data Science

APHRC, Summer 2015



Outline

Short Paper
Title

Christensen

Introduction

Replication

Project Protocol,
Reporting Standards

Workflow

Data Sharing

1 Introduction

2 Replication



Reproducibility & Transparency

Short Paper
Title

Christensen

Introduction

Replication

Project Protocol,
Reporting Standards

Workflow

Data Sharing

- What are problems associated with reproducibility?
- What are solutions to these problems?
- What are practical tools to implement these solutions?



Replication

Short Paper
Title

Christensen

Introduction

Replication

Project Protocol,
Reporting Standards

Workflow

Data Sharing

- 1 The Problem (JMCB Project)
- 2 Project Protocol, Reporting Standards
- 3 Organizing Workflow
- 4 Code & Data Sharing



Project Protocol, Reporting Standards

Short Paper
Title

Christensen

Introduction

Replication

Project Protocol,
Reporting Standards

Workflow

Data Sharing

Make sure you report everything another researcher would need to replicate your research.

- Find the appropriate reporting standard for your field and follow it: <http://www.equator-network.org/>
- Report the nuts and bolts of the project implementation in a detailed protocol:
<http://www.spirit-statement.org>



Workflow

Short Paper
Title

Christensen

Introduction

Replication

Project Protocol,
Reporting Standards

Workflow

Data Sharing

“Reproducibility is just collaboration with people you don’t know, including yourself next week”
—Philip Stark, UC Berkeley Statistics

Practical coding and organizational suggestions

- Long (2008) *The Workflow of Data Analysis Using Stata*
 - Making any changes to a file that has been posted/shared means it gets a new name.
 - Use version commands to ensure others get same results.
- Literate programming (extensive commenting, making the aim of code reading by a human)
- R Markdown, integration of analysis and output.



Data Sharing

Short Paper
Title

Christensen

Introduction

Replication

Project Protocol,
Reporting Standards

Workflow

Data Sharing

Post your code and your data in a trusted public repository.

- Find the appropriate repository:
<http://www.re3data.org/>
- Repositories will last longer than your own website.
- Repositories are more easily searchable by other researchers.
- Repositories will store your data in a non-proprietary format that won't become obsolete.