

CURATION AND VERIFICATION OF REPLICATION DATA FOR SOCIAL SCIENCE JOURNALS

THOMAS M. CARSEY, DIRECTOR

carsey@unc.edu



UNC
THE ODUM INSTITUTE

THE ODUM INSTITUTE FOR RESEARCH IN SOCIAL SCIENCE
228 DAVIS LIBRARY, CB# 3355
UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL
WWW.ODUM.UNC.EDU

THE ODUM INSTITUTE



Domain knowledge and statistics expertise



Established social science data archive



Commitment to research transparency and openness

JOURNAL POLICY



AJPS Replication and Verification Policy

The corresponding author of a manuscript that is accepted for publication in the *American Journal of Political Science* must provide replication materials that are sufficient to enable interested researchers to reproduce all of the analytic results that are reported in the text and supporting materials.

When the final draft of the manuscript is submitted, the replication materials will be verified to confirm that they do, in fact, reproduce the analytic results reported in the article.

Jacoby, W., & Lupton, R. (2015). AJPS Replication and verification policy. Retrieved from <https://ajps.org/ajps-replication-policy/>

JOURNAL POLICY

SUPPORTING DOCUMENTATION & INFORMATION

- ➔ README.txt file containing names of all files and information on how to replicate findings
- ➔ Codebook with variable definitions and value labels
- ➔ Software version information
- ➔ Citations for source datasets

ANALYSIS DATASET(S)

- ➔ Analysis dataset file(s) in a file format readily accessible to the social science research community
- ➔ Software command file(s) for reconstructing the analysis dataset from the original data source(s)

COMMAND FILES & TOOLS

- ➔ Commands needed to reproduce all tables and figures in the article and supplementary materials
- ➔ Commands/instructions for installing macros or packages
- ➔ Comment statements used to explain analysis steps and distinguish commands for tables and figures

JOURNAL WORKFLOW

- Journals already have an existing workflow
- Roles include: authors, editors, assistant editors, reviewers, production team, publisher
- Complying with a data replication and/or verification policy adds another layer to the workflow
- The challenge is to do so efficiently and effectively
- This involves deciding who performs replication and/or verification tasks, when those tasks occur in the overall workflow, and expectations for communication and reporting

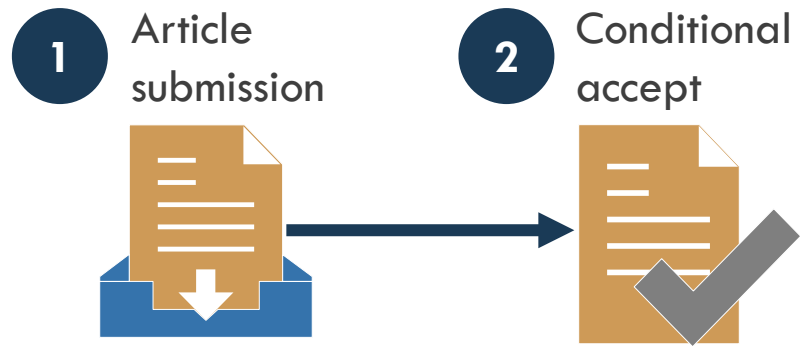
DATA VERIFICATION

DATA VERIFICATION

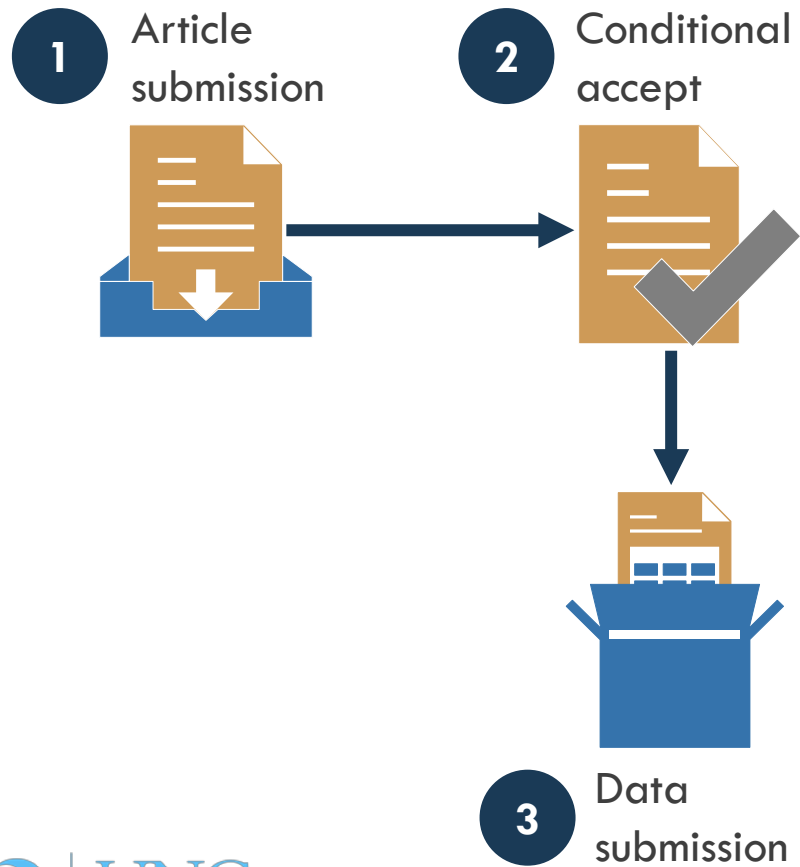
1 Article submission



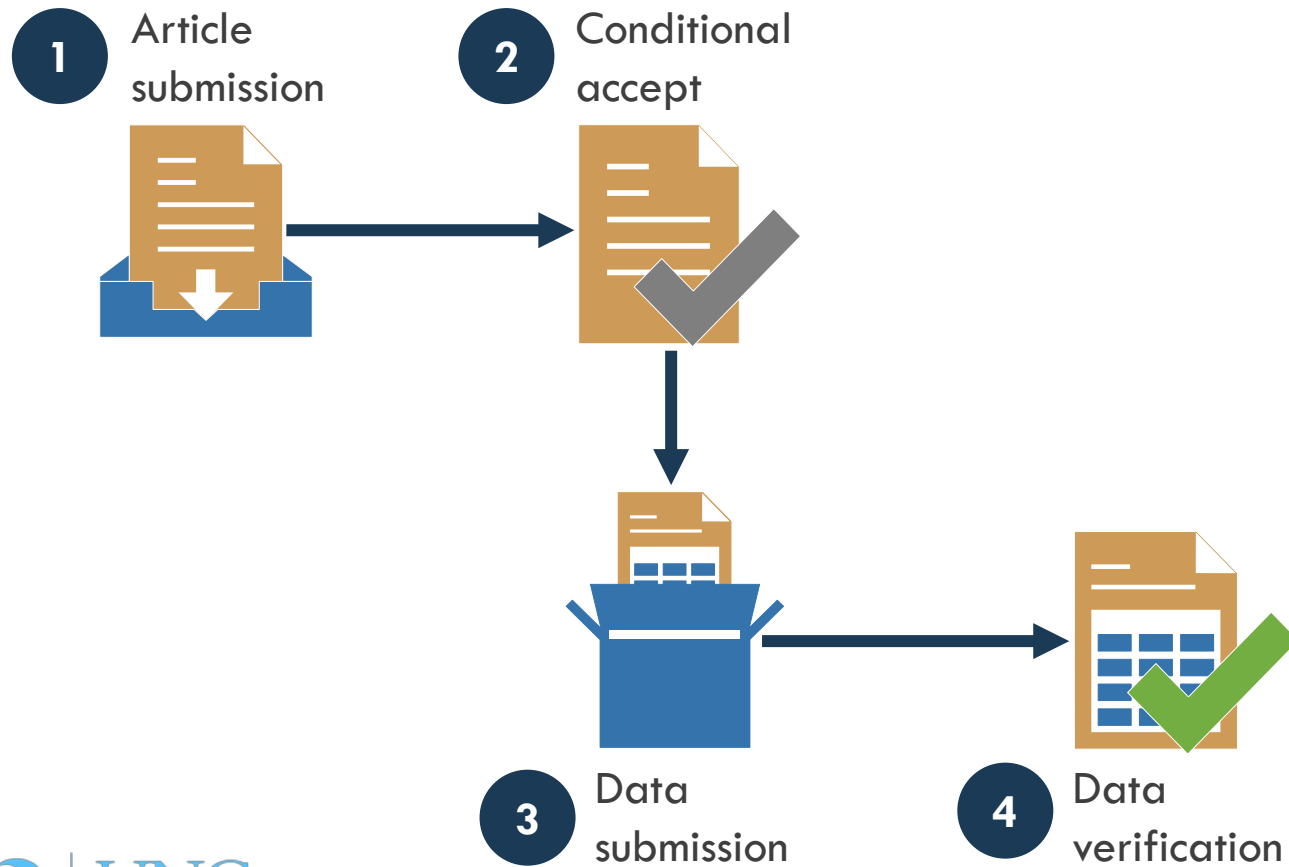
DATA VERIFICATION



DATA VERIFICATION



DATA VERIFICATION



DATA VERIFICATION

DATA CURATION



REPLICATION VERIFICATION

DATA VERIFICATION

DATA CURATION

- ✓ Review materials for completeness
- ✓ Identify confidentiality risks
- ✓ Identify incomplete, inconsistent or missing variable/value labels
- ✓ Assess file formats for suitability for long-term preservation
- ✓ Enhance descriptive metadata



REPLICATION VERIFICATION

DATA VERIFICATION

DATA CURATION

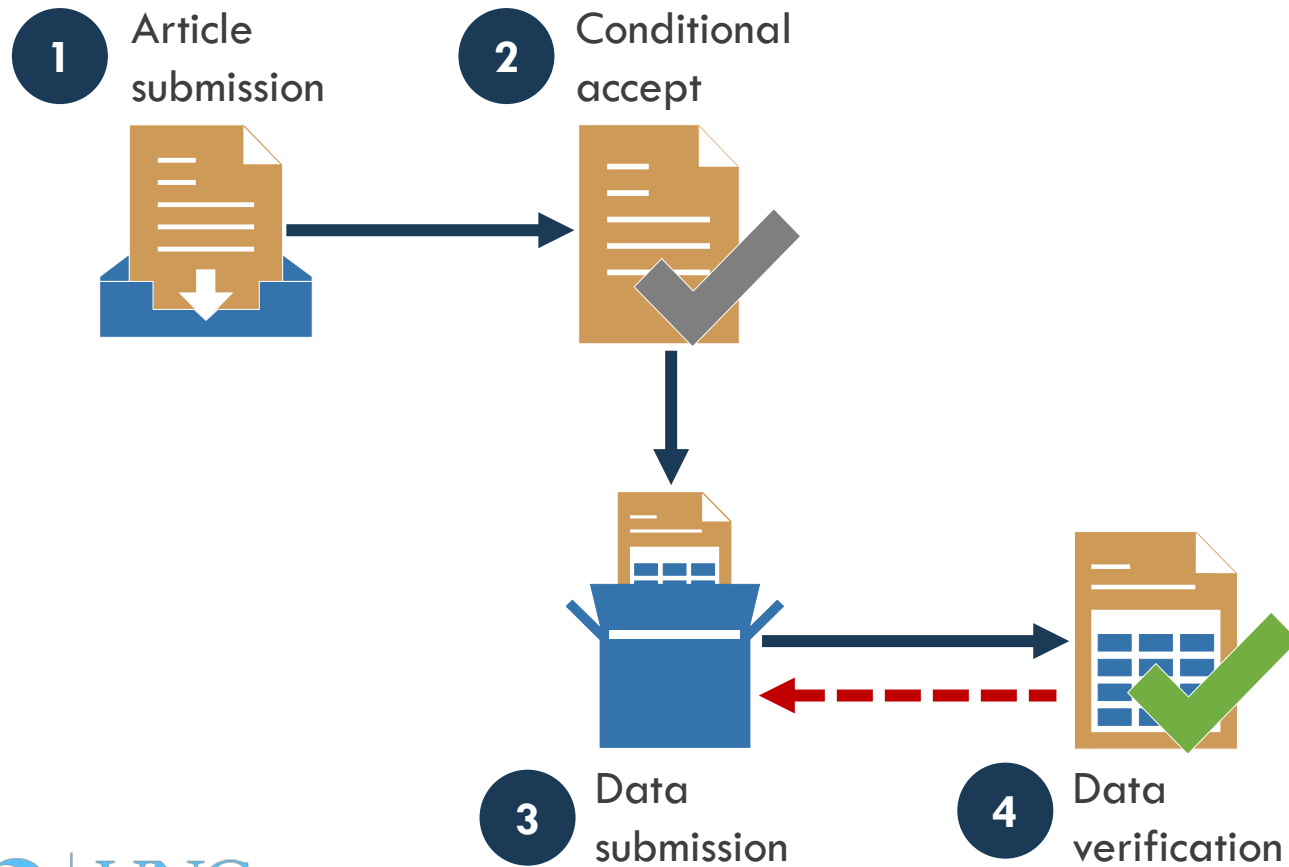
- ✓ Review materials for completeness
- ✓ Identify confidentiality risks
- ✓ Identify incomplete, inconsistent or missing variable/value labels
- ✓ Assess file formats for suitability for long-term preservation
- ✓ Enhance descriptive metadata



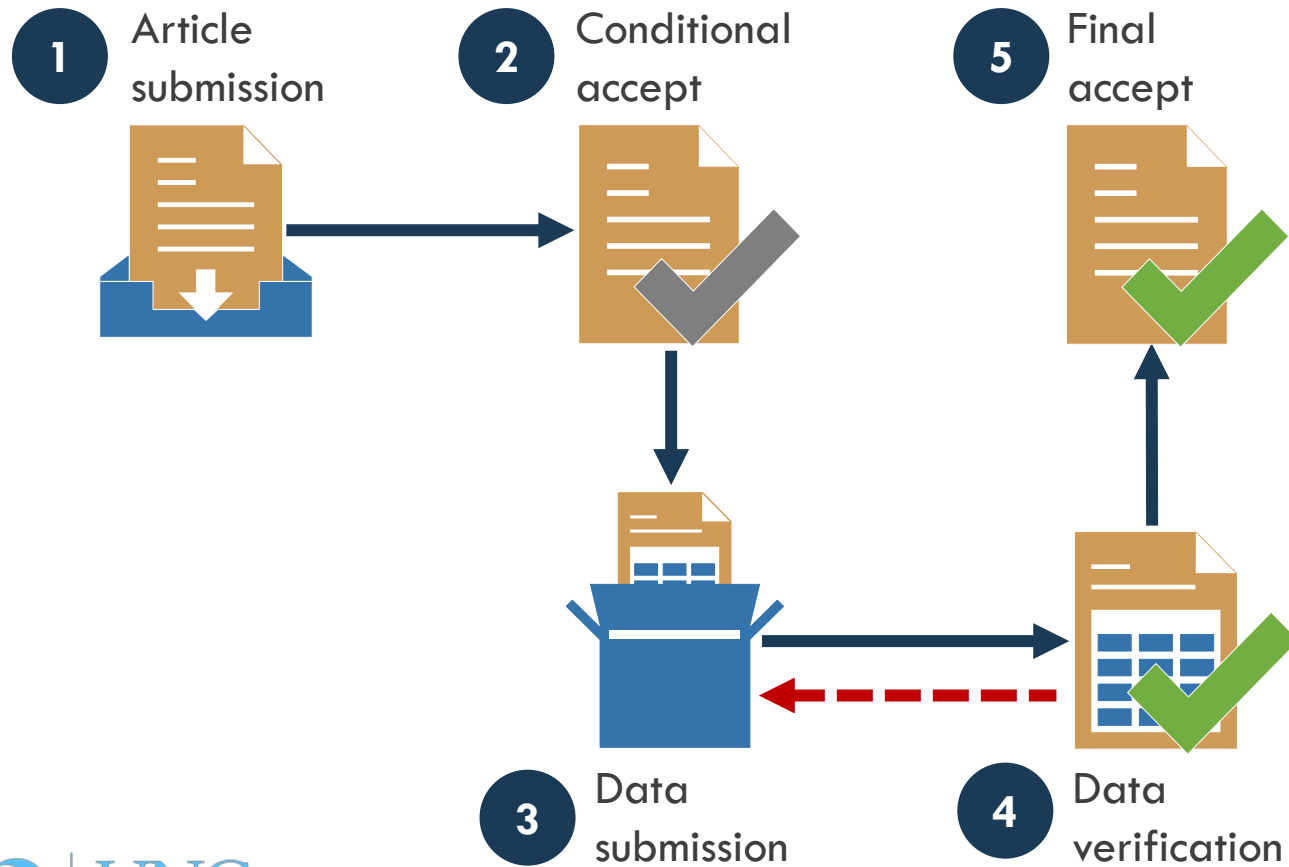
REPLICATION VERIFICATION

- ✓ Review code for inclusion of commands and comments required for proper execution
- ✓ Execute code
- ✓ Compare outputs to tables and figures in the manuscript

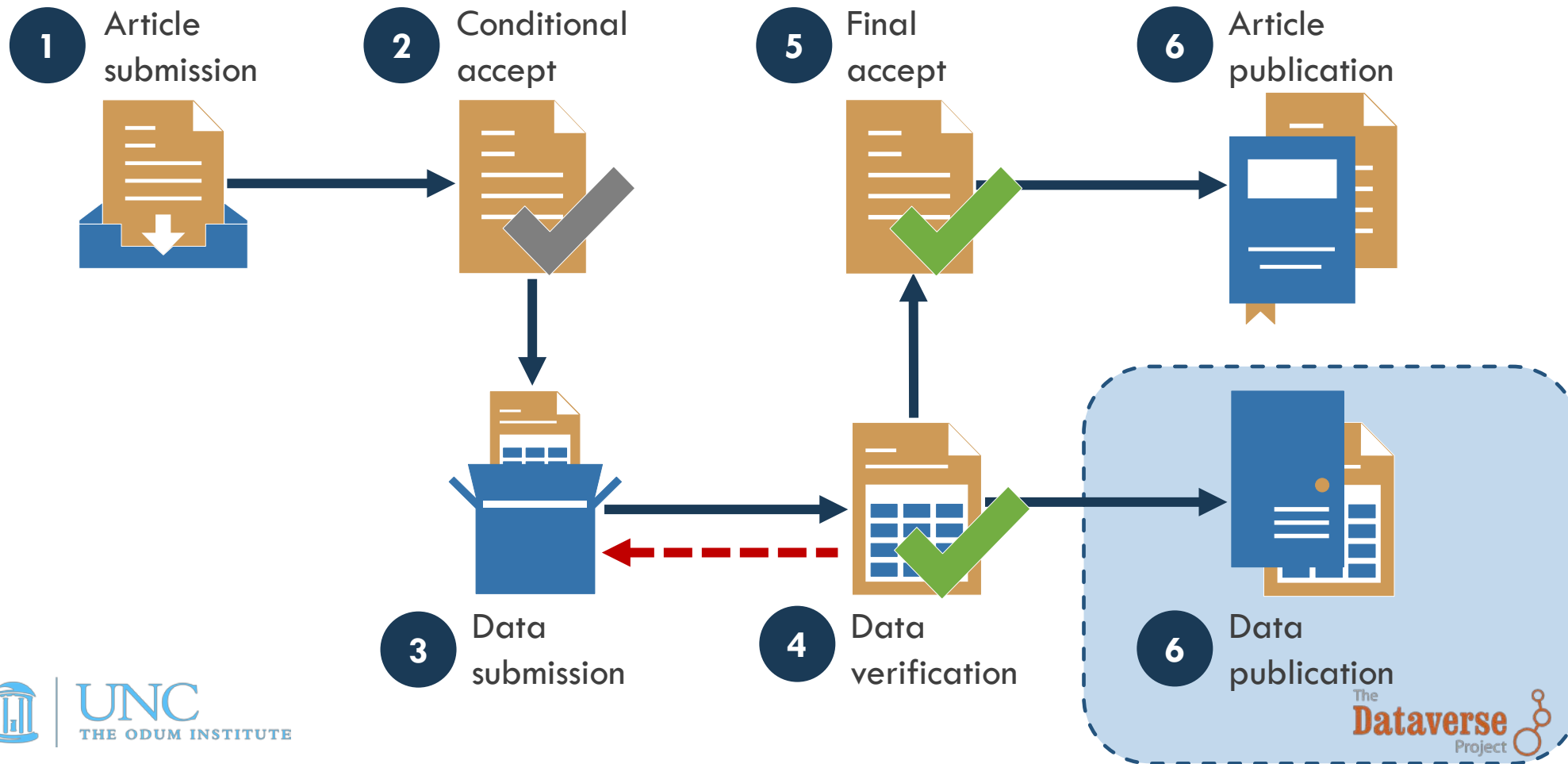
DATA VERIFICATION



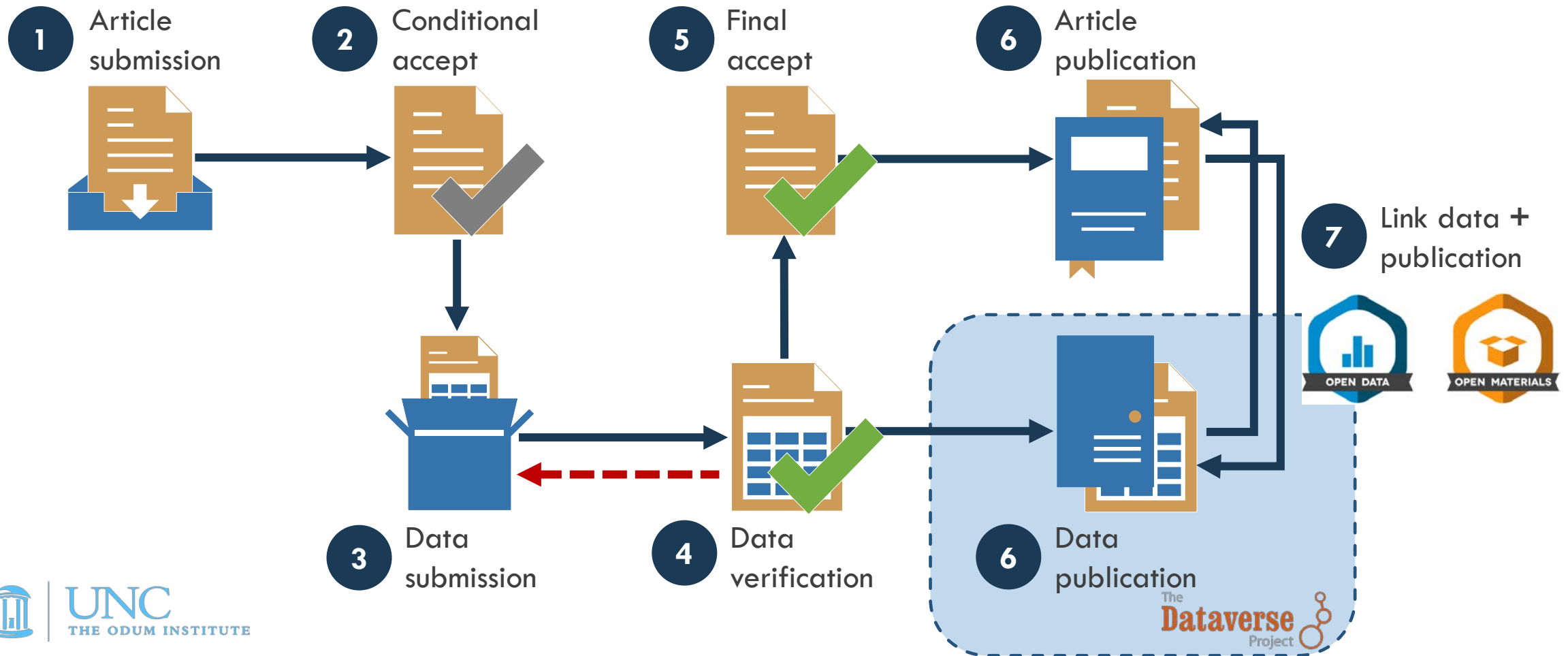
DATA VERIFICATION



DATA VERIFICATION



DATA VERIFICATION



AJPS GUIDELINES AND CHECKLISTS

- AJPS website ajps.org
 - Guidelines for accepted articles
- AJPS Workplan ([Here](#))
- Manuscript Verification Report ([Here](#))

METRICS (MARCH 2015-MAY 2016)

METRICS (MARCH 2015-MAY 2016)



66

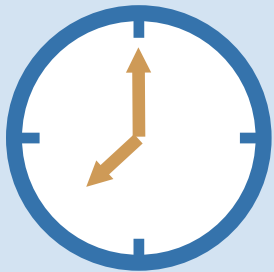
manuscripts completed

METRICS (MARCH 2015-MAY 2016)



66

manuscripts completed



8

hours (average)

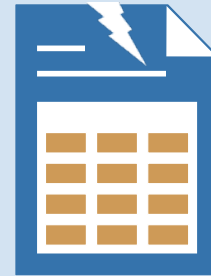


METRICS (MARCH 2015-MAY 2016)



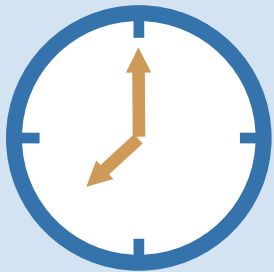
66

manuscripts completed



92%

require resubmission



8

hours (average)

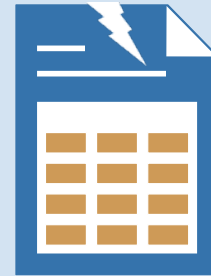


METRICS (MARCH 2015-MAY 2016)



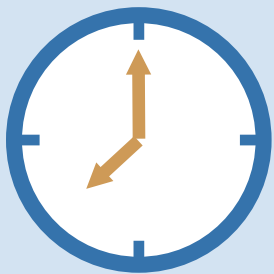
66

manuscripts completed



92%

require resubmission



8

hours (average)

- ✗ Missing or incomplete documentation
- ✗ Missing or non-documented code
- ✗ Code execution errors
- ✗ Rounding errors
- ✗ Mismatched figures or tables
- ✗ Incomplete data source citations



LESSONS LEARNED

- The value of expertise
- Cooperation between author and editor and archive is critical
- Improved training and tools to help authors adopt best practices from the start will pay off big at the end
- Scalability and resources – it takes resources to engage in data replication and verification, but efficiencies can be gained by scaling up

NEXT STEPS

- Investigate **automation and integration opportunities** to streamline workflows and optimize efficiency
- **Develop new tools** to assist researchers in creating high quality replication data packages
- Work with the community to **support open science & reproducible research**

Data Curation & Verification Service

The **Odum Institute** provides a third-party data curation and verification service that empowers journals to enforce their data sharing and replication policies, which includes:

Data Curation

Review replication materials to ensure they are sufficient for **independent use**

Perform **data quality assurance** checks for missing or mislabeled variables & values, confidentiality risks, & suitable file formats

Link data to published articles using persistent identifiers (DOI)

Verification

Review analysis code for the inclusion of commands & comments required for execution

Verify **successful execution** of analysis code & confirm outputs match tables & figures in the manuscript

Provide a **standardized report** to communicate any identified issues

The level of curation and verification can be customized based on journal needs and available resources

Data Curation & Verification Benefits

Demonstrates a commitment to transparent & reproducible research

Ensures authors comply with established journal data policies

Guarantees data are properly curated & preserved within a trusted repository

Establishes a stable bidirectional link between data & published articles

Enhances the visibility of published articles through independent discovery of associated data

Integrate data quality review **within** the current manuscript publication workflow

THE ODUM INSTITUTE
FOR RESEARCH IN SOCIAL SCIENCE



THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL

For more information please contact odumarchive@unc.edu & visit us at www.odum.unc.edu

CONNECT WITH THE ODUM INSTITUTE

Acknowledgments:

William Jacoby,

Editor of AJPS

Thu-Mai Christian

tlchristian@unc.edu

Sophia Lafferty-Hess

slaffer@email.unc.edu

