

ADAPTING DATA VERIFICATION WORKFLOWS TO ACCOMMODATE RESTRICTED REPLICATION DATA

THU-MAI CHRISTIAN | ASSISTANT DIRECTOR FOR ARCHIVES



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 H. W. ODUM INSTITUTE FOR RESEARCH IN SOCIAL SCIENCE

- Founded in 1924 by Howard W. Odum
- Oldest university-based interdisciplinary social science research institute in the U.S.
- To help grow and lead a world-class social science research infrastructure at the University of North Carolina at Chapel Hill to ensure that researchers can conduct scientifically rigorous research that contributes to better lives of the citizens of North Carolina and the World...



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



SPPQ Guidelines for Preparing Replication Files

State Politics and Policy Quarterly (SPPQ) requires the authors of all accepted manuscripts to **provide replication files** before the article enters the production state of the publication process.

The purpose of replication materials is to **enable interested researchers to reproduce all of the analyses** that are reported and presented in articles that appear in *SPPQ*. Authors should always consider this basic objective when preparing replication materials.

Bonneau, C., & Kanthak, K. (2015). Guidelines for preparing replication files. Retrieved from http://spa.sagepub.com/site/misc/SPPQ_Instructions/guidelines_for_preparing_replication_files.pdf

JOURNAL POLICY

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

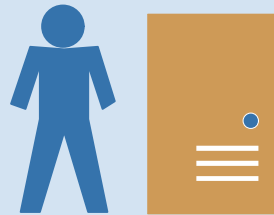
SUPPORTING DOCUMENTATION & INFORMATION

- ➔ 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

THE ODUM INSTITUTE



Domain knowledge and statistics expertise



Established social science data archive

OPENDAT



Commitment to research transparency and openness



UNC
THE ODUM INSTITUTE

DATA VERIFICATION

DATA CURATION

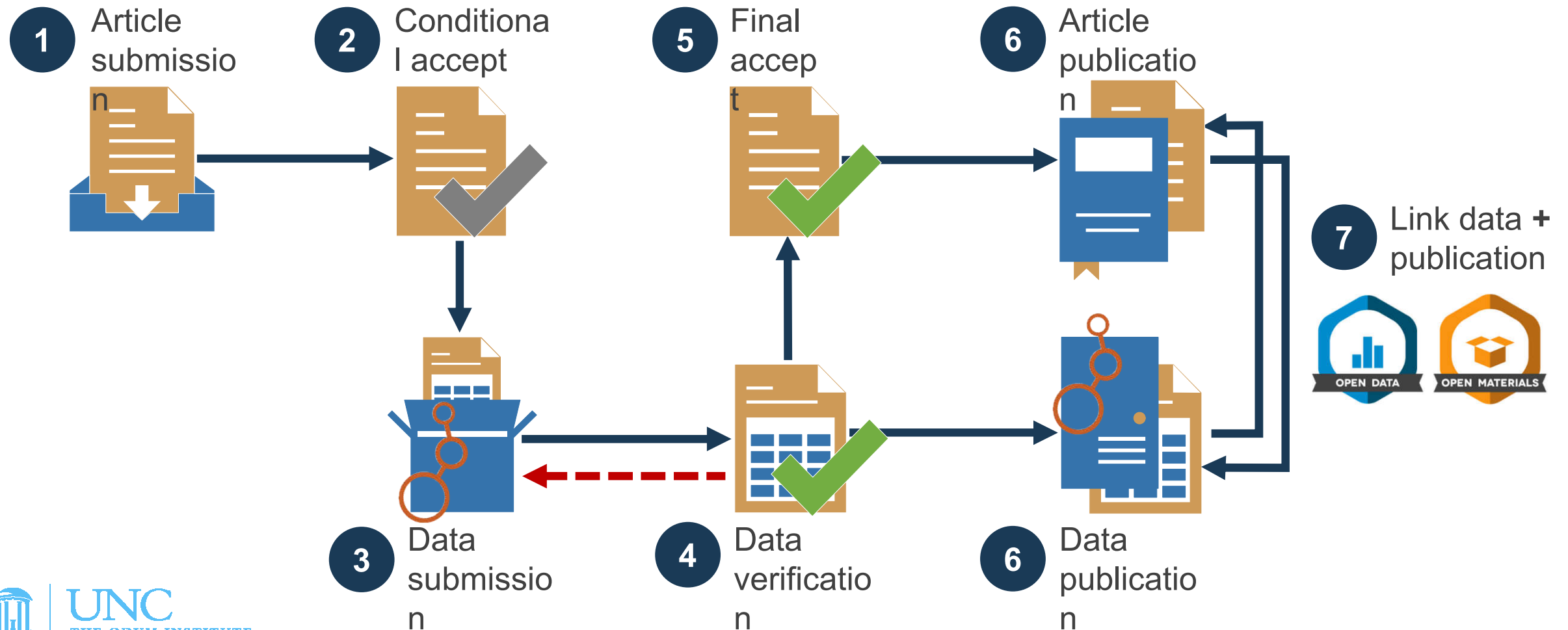
- ✓ Review materials for completeness
- ✓ Identify confidentiality risks
- ✓ Identify incomplete, inconsistent or missing variable/value labels
- ✓ Assess file formats for suitability for long-term preservation
- ✓ Enhance descriptive metadata
- ✓ Link data to published article



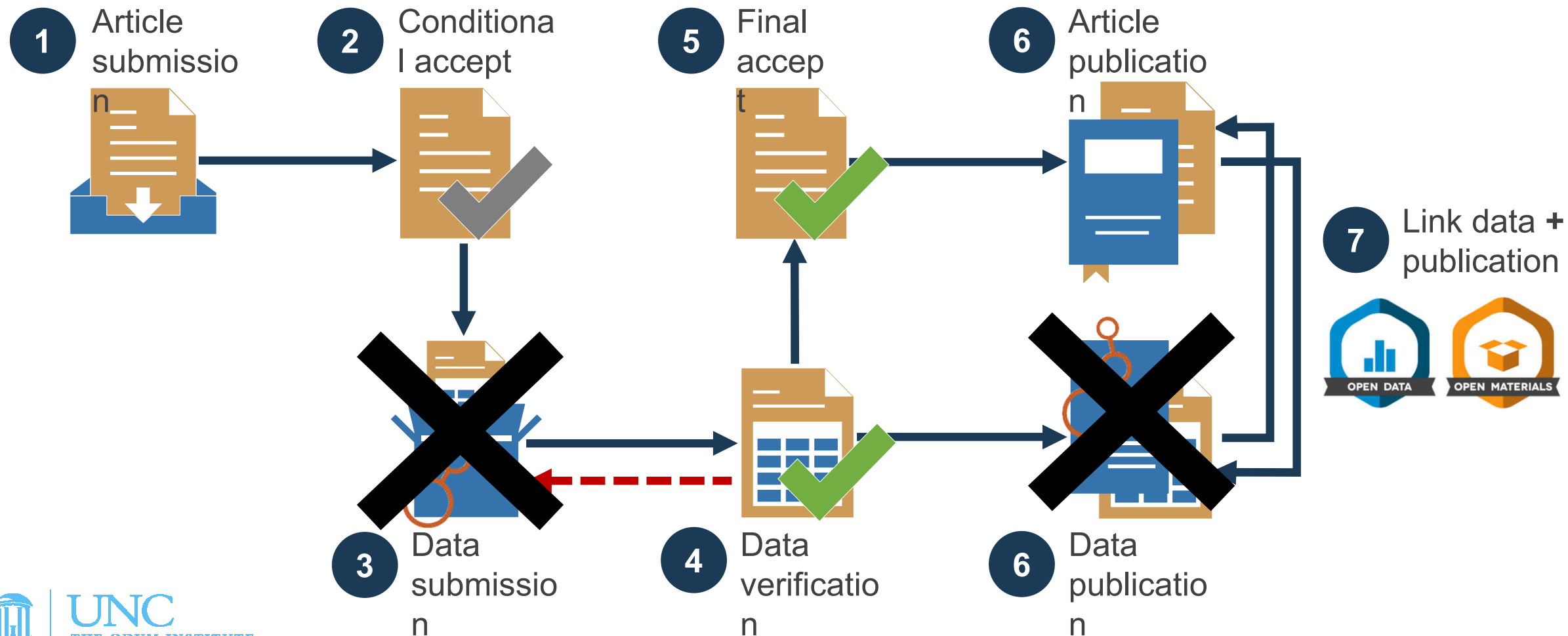
REPLICATION VERIFICATION

- ✓ 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 UNDER CONSTRAINT

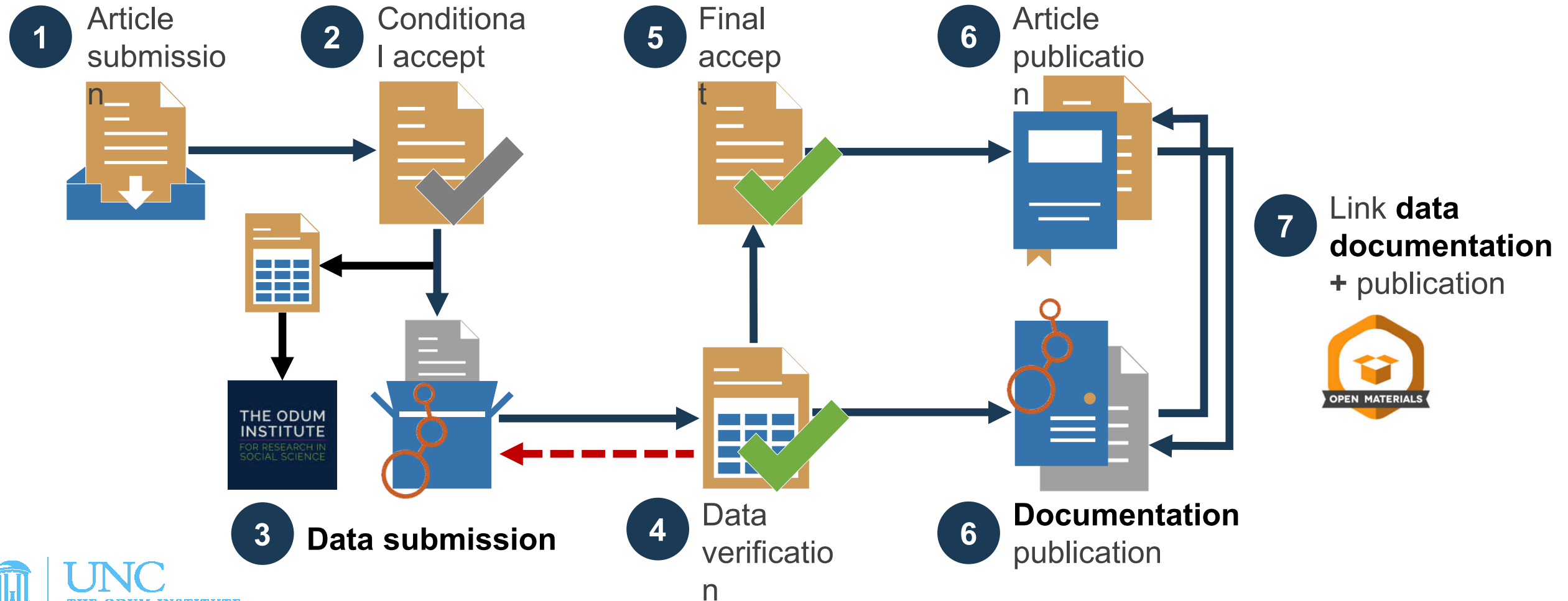


DATA VERIFICATION – DATA UNDER CONSTRAINT



- Terms of **license/copyright** prohibit redistribution of data
Author obtains permission from data provider to submit data to the Odum Institute for the purpose of verification.
- Data containing sensitive information **require security measures**
Author submits data directly to the Odum Institute, which implements all necessary security measures to protect the data.
- Account **registration and approval** is required to access data
Odum Institute registers with data provider to obtain approval to access data for the purpose of verification.

DATA VERIFICATION – DATA UNDER CONSTRAINT



DATA VERIFICATION – DATA UNDER CONSTRAINT



8%
under constraint

- Sensitive/proprietary data deleted upon successful verification
- Data requested directly from the data provider
- Author paid fee to allow access to data for verification

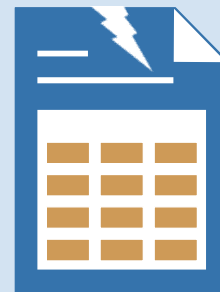
DATA VERIFICATION – DATA UNDER CONSTRAINT



8%

under constraint

- Sensitive/proprietary data deleted upon successful verification
- Data requested directly from the data provider
- Author paid fee to allow access to data for verification



>90%

require resubmission

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

DATA VERIFICATION – DATA UNDER CONSTRAINT

- Develop new technologies to facilitate the secure exchange of proprietary and sensitive data within manuscript publication and data verification workflows.
- Establish standard systems in consultation with IRBs that support compliance with rules and regulations for human subjects research during verification processes.
- Cultivate relationships with data providers to encourage their participation in data verification processes.
- **Provide education, tools, and services to authors to assist them with the production of high-quality datasets.**

CONNECT WITH THE ODUM INSTITUTE

Thu-Mai Christian
tlchristian@unc.edu

