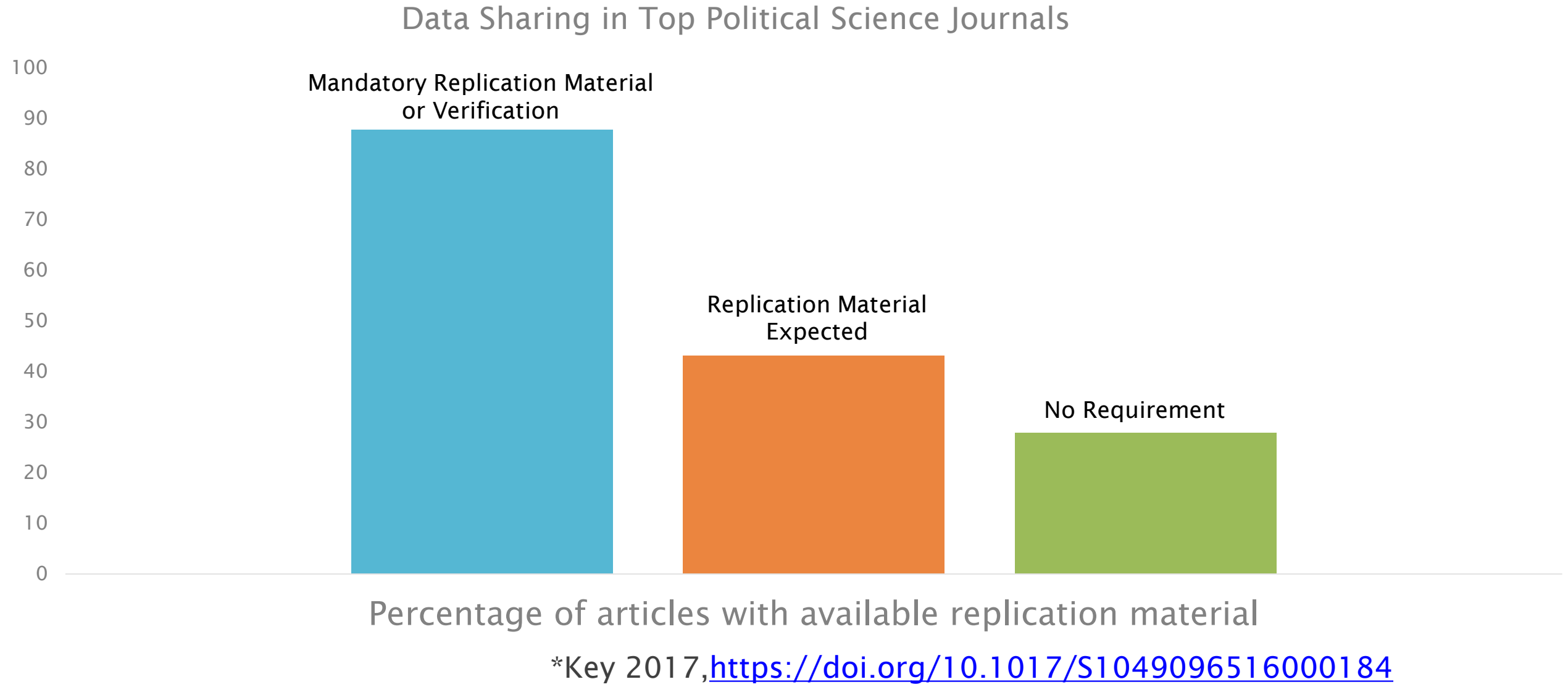
Qualitative and Multi-Method Badging and Verification

Colin Elman (QDR)
Sebastian Karcher (QDR)
Data-PASS Workshop, APSA 2019
August 28, 2019

QDR
QUALITATIVE DATA REPOSITORY

Sharing Data and Code: Requirements Work



*Key 2017,https://doi.org/10.1017/S1049096516000184

Are There Other Ways to Encourage Sharing?

- Requirements may be undesirable / too burdensome
- Even with requirements, additional incentives might help
- Some evidence that pure requirements lead to low-quality data deposits

How about… gold stars for grown-ups: Open Science Badges

Replication Materials

The data, code, and any additional materials required to replicate all analyses in this article are available on the *American Journal of Political Science* Dataverse within the Harvard Dataverse Network, at: https://doi.org/10.7910/DVN/HI1X4O.

This Can't Possibly Work?

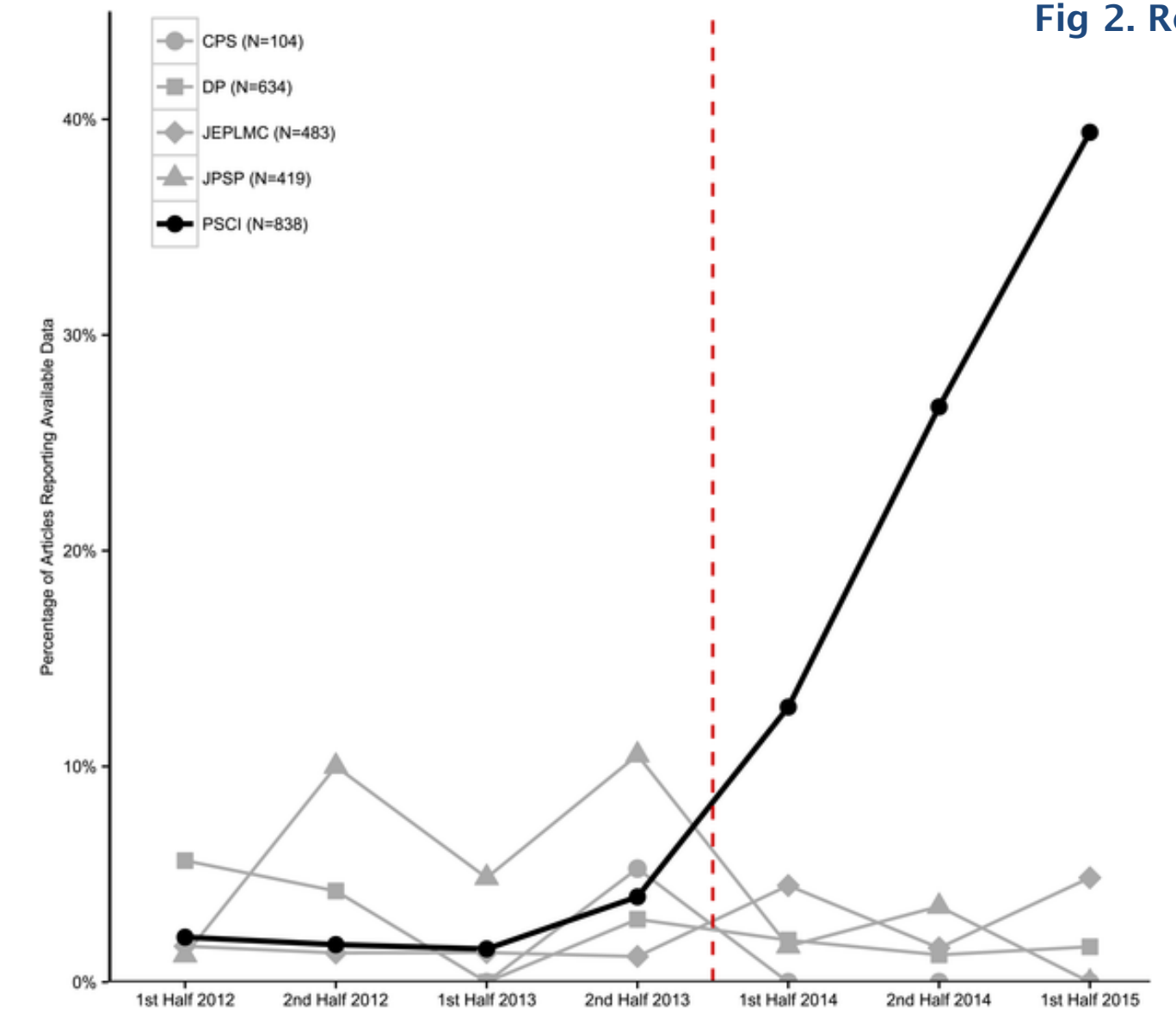


Fig 2. Reportedly available data.

Kidwell MC, Lazarević LB, Baranski E, Hardwicke TE, Piechowski S, et al. (2016) Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency. PLOS Biology 14(5): e1002456. https://doi.org/10.1371/journal.pbio.1002456

Who Do I Have to Pay to Use These?

- No one!
- Open Science Badges are created by the Center for Open Science (COS)

"The badges are free to use with attribution of their source. Download the images and get started now: https://osf.io/tvyxz/files/"

What Do They Mean?

The Open Data badge is awarded when digitally-shareable data necessary to reproduce the reported results are publicly available.

The Open Materials badge is earned by making publicly available the components of the research methodology needed to reproduce the reported procedure and analysis.

The Preregistered badge is earned for preregistering research.

All Badge descriptions from https://osf.io/sxn73/

How About Sensitive Data?

- Data that can only be accessed under special conditions ("protected access") can receive a dedicated "PA" Open Data Badge

Who Checks This?

The regular badges work on a system of social trust based on the authors' disclosure:

"Disclosure requires authors to provide public statements affirming achievement of badge criteria. The certifying organization evaluates the disclosure before issuing the badge, but *does not do more than a cursory evaluation of the data, materials, or registration*."

Peer Reviewed Open Badges

Four options

1. Reviewers of the report can also review associated data, materials, and preregistration.
2. Additional reviewer(s) can be recruited specifically for badge review.
3. An organization staff member could provide badge review.
4. An independent organization could provide badge review as a service for the certifying organization.

Verification as a subset of Reappraisal

- Unsettled semantic field. Too many terms being used inconsistently.
- Several different typologies and unlikely to be unified any time soon.
- For purposes of this presentation, apply recent typology developed by Gerring.
- Umbrella concept of “reappraisal”: follow-up studies that bear on the truth-value of an initial study
- Verification is a subset of reappraisal, involving same data and same analysis.

Gerring: Reappraisals

		MEASUREMENT/ANALYSIS:	
		Same	*Different*
EVIDENCE:	*Same data*	1. Verification	2. Sensitivity
	Same population	3. Reproduction	4. Reanalysis
	Different population	5. Extension	

From John Gerring, "Coordinating Reappraisals" in Colin Elman, John Gerring, and James Mahoney, Eds., *The Production of Knowledge: Enhancing Progress in Social Science* (CUP, forthcoming)

Mapping other terms to Gerring's typology: Reproducibility and Replicability

"We define *reproducibility* to mean computational reproducibility—obtaining consistent computational results using the same input data, computational steps, methods, and code, and conditions of analysis; and *replicability* to mean obtaining consistent results across studies aimed at answering the same scientific question, each of which has obtained its own data. In short, reproducibility involves the original data and code; replicability involves new data collection and similar methods used by previous studies."

National Academies of Sciences, Engineering, and Medicine 2019. *Reproducibility and Replicability in Science*. Washington, DC: The National Academies Press. https://doi.org/10.17226/25303.

Qualitative Verification

- Gerring's approach explicitly focuses on quantitative research, but acknowledges it could be adapted for qualitative.
- QDR sees qualitative verification as having similar sensibilities and aspirations as computational reproduction: ascertaining whether same data and same analysis support results
- But differences in epistemology/method and material processes require more nuanced and flexible approach
- Not as precise – not numeric results of a calculation
- Not as deterministic – not product of an explicit algorithm
- Correspondingly, qualitative reappraisals don't generate (in Gerring's terms) *exact* corroborations

Gerring: Levels of Corroboration

(a)	*Not possible*	• Theory unfalsifiable, data unavailable, or procedures unclear
(b)	*No corroboration*	• Reappraisal estimate falls outside confidence interval of the original estimate
(c)	*Weak corroboration*	• Reappraisal estimate falls inside confidence interval of the original estimate in some tests
(d)	*Strong corroboration*	• Reappraisal estimate falls inside the confidence interval of the original estimate
(e)	*Exact corroboration*	• Estimate is exactly reproduced

From John Gerring, "Coordinating Reappraisals" in Colin Elman, John Gerring, and James Mahoney, Eds., *The Production of Knowledge: Enhancing Progress in Social Science* (CUP, forthcoming)

Gerring: Validation Types

REAPPRAISAL TYPES	CORROBORATION TYPES				
	(a) Not possible	(b) No corroboration	(c) Weak corroboration	(d) Strong corroboration	(e) Exact corroboration
1. Same data, Same measurement/analysis	1(a)	1(b)	1(c)	1(d)	1(e)
2. Same data, Different measurement/analysis	2(a)	2(b)	2(c)	2(d)	2(e)
3. Same population, Same measurement/analysis	3(a)	3(b)	3(c)	3(d)	3(e)
4. Same population, Different measurement/analysis	4(a)	4(b)	4(c)	4(d)	4(e)
5. Different population, Same measurement/analysis	5(a)	5(b)	5(c)	5(d)	5(e)

From John Gerring, "Coordinating Reappraisals" in Colin Elman, John Gerring, and James Mahoney, Eds., *The Production of Knowledge: Enhancing Progress in Social Science* (CUP, forthcoming)

Other contrasts/similarities between qualitative verification and computational reproduction:

- More diverse. Quantitative tools, discrete set(s) of techniques, widely shared understandings held in common. For qualitative, depends on the nature of the data and analysis.
- Less bounded. Computational reproduction, focuses on numbers in the tables. Qualitative, much more porous boundary between parts of article that are analogous to the tables and to subsequent discussion.
- Both try to avoid substantive questions that are more properly subject of journal's review of the article.

Other contrasts/similarities between qualitative verification and computational reproduction:

- Both have difficult decisions to make about what/how much supplemental material to verify (e.g. article appendices, supplements)
- Both need to ascertain whether/ which badges to issue: production and analytic transparency map to open materials badge, data access to open data badge
- Both need to cope with data that are constrained - which in some circumstances will make verification impossible, and in others possible but will impact whether/which badges can be issued

Quantitative 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

+

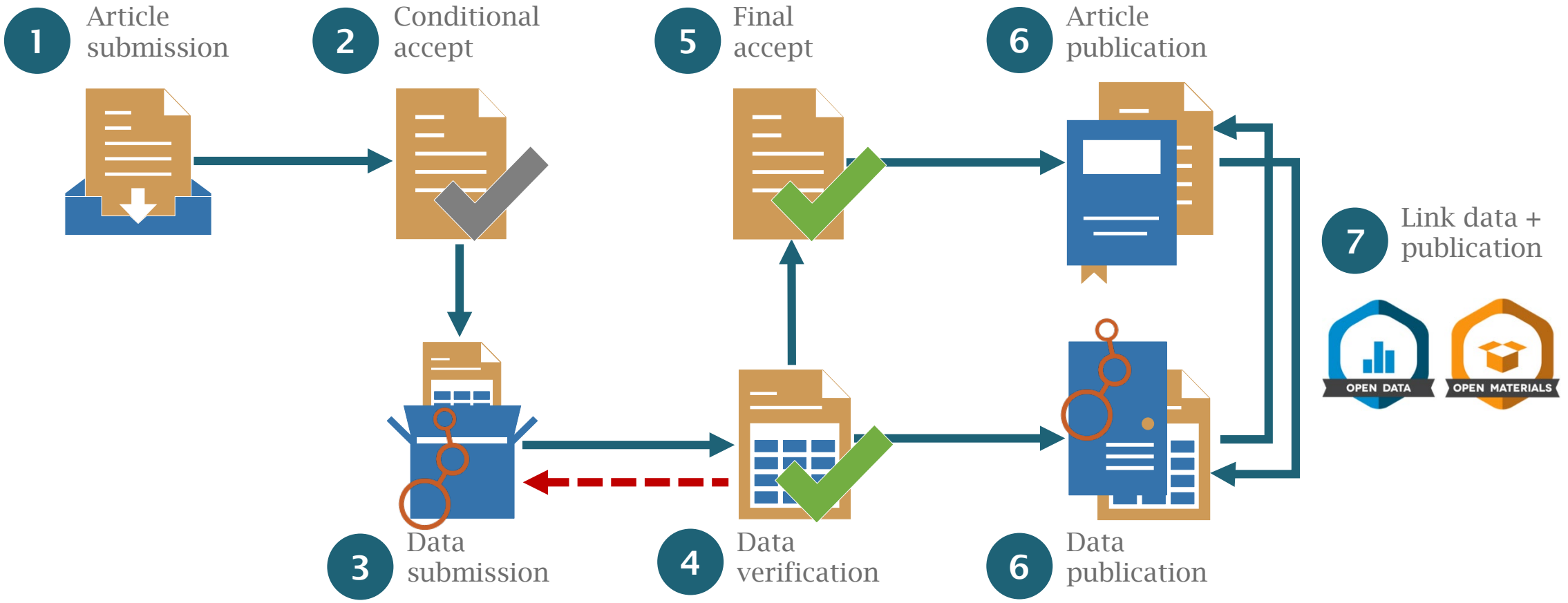
COMPUTATIONAL REPRODUCIBILITY

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

Slide adapted from: Thu-Mai Christian, "Adapting data verification workflows to accommodate restricted replication data," Data-PASS workshop 2016

Quantitative data verification

Slide adapted from: Thu-Mai Christian, "Adapting data verification workflows to accommodate restricted replication data," Data-PASS workshop 2016



Qualitative data verification

DATA CURATION

- Review materials for completeness (note that qualitative materials will be different, e.g. interview appendix)
- Identify confidentiality risks
- Review file names for consistency
- Assess file formats for suitability for long-term preservation
- Enhance descriptive metadata
- Link data to published article

+

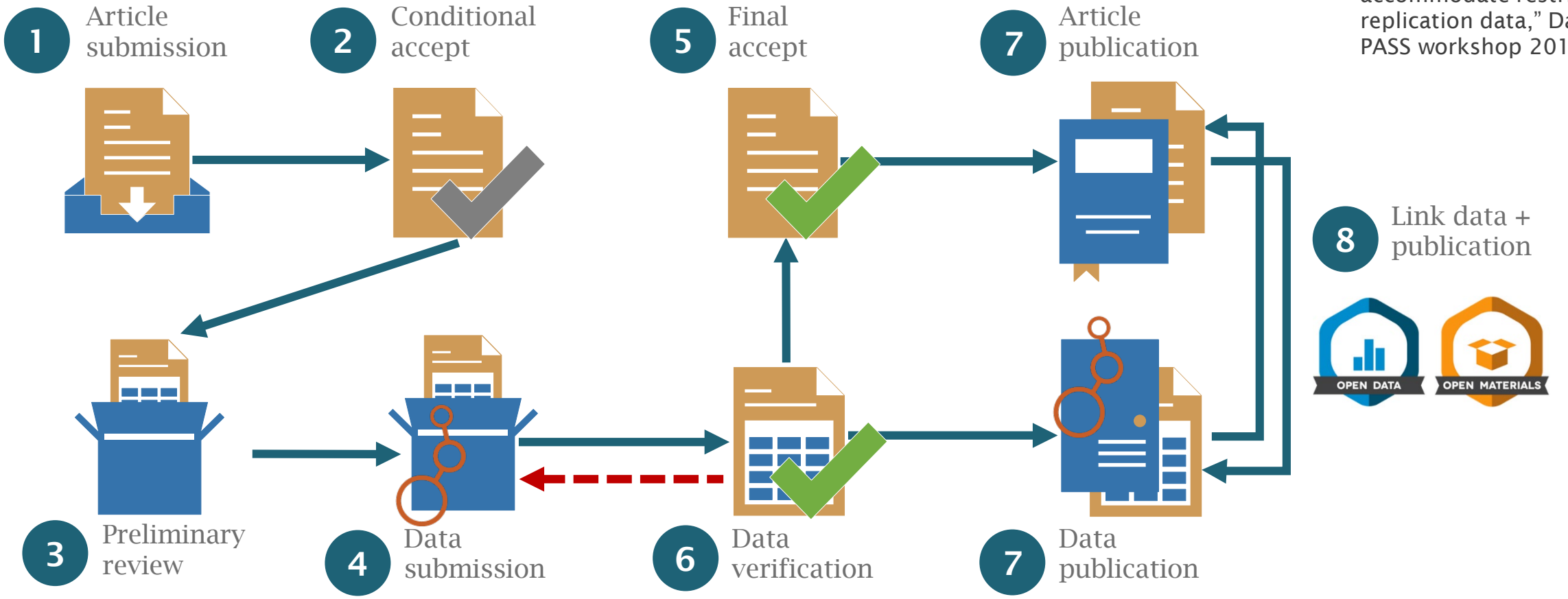
VERIFICATION

- Context dependent review of data and analysis. Examples:
  - Whether cited sources support entries in a nominal or ordinal typology
  - Whether interview quotations in text are representative of larger set
  - Whether cited sources support central elements of a case narrative

Slide adapted from: Thu-Mai Christian, "Adapting data verification workflows to accommodate restricted replication data," Data-PASS workshop 2016

Qualitative data verification

Slide adapted from: Thu-Mai Christian, "Adapting data verification workflows to accommodate restricted replication data," Data-PASS workshop 2016



QDR's Qualitative Verifications for the *AJPS*

- Three verifications completed.
- One pre-submission assessment
- One manuscript declined for verification
- For each of the three verifications, QDR looked three times: initial assessment, main report, subsequent sign-off.
- Collectively, three verifications included data from interviews, archival sources, and secondary sources.
- Working with Odum to coordinate verifications that have significant qualitative and quantitative components

Example: Carnegie and Carson

- Usually would not discuss a particular verification, but authors wrote a blog about their experience.
- Authors used a variety of qualitative materials including archival documents, expert interviews, and other primary and secondary sources.
- Authors uploaded the raw archival material they used, along with the relevant interview excerpts, to the AJPS Dataverse, which AJPS then gave QDR access.
- A few other secondary sources cited by the authors were obtained by QDR, and cited pages were scanned and saved as pdf files.

Carnegie and Carson (cont'd)

- QDR did not have any direct interactions with the authors. We reported to AJPS, which then communicated with the authors.
- For each claim we evaluated, we assessed the prima facie congruence between the claim and the evidence in the cited source(s). QDR used three codings:
  - (1)NOT SUPPORTED, indicating that the claims and the evidence provided did not match;
  - (2) PARTIALLY SUPPORTED, indicating that some but not all claims in the statement matched with the evidence provided; and
  - (3) SUPPORTED, indicating a match between claim and evidence.
-

Carnegie and Carson (cont'd)

- In our report to AJPS, for each claim, we provided the following information:

  (1) excerpt of statement in manuscript;

  (2) page number of excerpt in manuscript;

  (3) source reference;

  (4) whether the source was provided as (a) citation, (b) URL, and/or (c) on AJPS Dataverse (DV);

  (5) the filename of the source (The secondary source files were saved according to the footnote or page number (if an in-text citation) where the reference was cited.

  (6) whether the statement was supported by the evidence provided;

  (7) if needed, an explanation for the conclusions drawn; and

  (8) any suggestions we had for how authors might respond.

Carnegie and Carson (cont'd)

- QDR also provided a separate list of unreferenced empirical statements that appear to warrant support.
- Statements we evaluated included claims that underpinned frequency counts in the authors' tables, and evidence-based arguments related to their theory's mechanisms.
- Our judgements on the referenced empirical sources were limited to their prima facie congruence with statements made by the authors. QDR did not express any opinions on the standing of those sources.
- QDR expressed no opinion on the authors' overall conclusions.

Carnegie and Carson (cont'd)

- The authors wrote in their blog that "preparing the materials for verification, and then responding to QDR's report, took roughly the same amount of time that verification processes of quantitative data typically do, so it did not delay the publication of our article."
- The authors also wrote that they found QDR's report to be "thorough, accurate, and helpful. While we had endeavored to support our claims fully in the original manuscript, we fell short of this goal on several counts, and followed each of QDR's excellent recommendations. Occasionally, this involved a bit more research, but typically this resulted in us clarifying statements, adding details, or otherwise improving our descriptions of, say, our coding decisions."

Carnegie and Carson (cont'd)

- "We believe that the verification both improved the quality of evidence and better aligned our claims with our evidence. Moreover, it increased our confidence that we had clearly and accurately communicated with readers. Finally, archiving our data will allow other scholars to access our sources and evaluate our claims for themselves, as well as potentially use these materials for future research. We thus came away with the view that qualitative transparency is achievable in a way that is friendly to researchers and can improve the quality of the work."