FROM POLICY TO PRACTICE: HOW JOURNAL-BASED DATA POLICIES ENCOURAGE SCIENTISTS’ ADOPTION OF REPRODUCIBLE RESEARCH PRACTICES

Thu-Mai Lewis

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the School of Information and Library Science

Chapel Hill
2023

Approved by:
Helen R. Tibbo
Colin Elman
Bradley Hemminger
Christopher A. Lee
Ayoung Yoon
ABSTRACT

Thu-Mai Lewis: From policy to practice: How journal-based data policies encourage scientists’ adoption of reproducible research practices
(Under the direction of Helen R. Tibbo)

According to several studies, researchers are not sharing the data underpinning their published scientific results, despite their general consensus that sharing data is critical to the research enterprise. Among other benefits, data sharing allows for verification of claims, which is essential to scientific integrity. Research funders, journal editors, and professional associations have insisted on the importance of data sharing by issuing policies and codes of ethics that mandate the practice. However, these mandates have not always been proven to compel researchers to share their data as evidenced by failed attempts to locate data underlying published results or sharing data that do not meet quality standards to allow for verification or reuse. This dissertation seeks to understand the incongruity between researchers’ belief that data sharing is essential to science and their failure to produce and share data underlying their reported findings—even when policy requires them to do so.

To address this phenomenon, the dissertation investigates the implementation and outcomes of the rigorous American Journal of Political Science (AJPS) verification policy that makes publication in the journal contingent on submission of data, code, and supporting documentation (i.e., the research compendium). Prior to publication, research compendia undergo a third-party verification process to confirm the computational reproducibility of findings presented in the manuscript. In most cases, authors fail initially to produce a
compendium that meets policy requirements for completeness, understandability, and computational reproducibility.

Using the theory of planned behavior (TPB) as a framework, the study investigates the specific behavioral factors that affect authors’ success or failure in producing reproducible research compendia. Employing a mixed-methods/grounded theory approach, the study analyzes records of verification results and interviews with AJPS authors who were subject to the verification policy to learn about their specific reproducible research practices (or lack thereof) and their outcomes. Based on the results of the study, I identify the most common and impactful issues that appear in submitted research compendia that render them non-reproducible, and suggest reasons that authors encounter these issues. Finally, I propose an extension of TPB that suggests how the policy compels and supports behaviors that promote research reproducibility.
To Ben August.
ACKNOWLEDGEMENTS

I never would have imagined that I would fulfill my mother’s wish of becoming a doctor. While not the type of doctor she was wishing for, I know that this achievement will make her proud. I have watched my mother hold herself to uncompromising standards for hard work even in the face of tragedy and hardship. It is only because I hold myself to her standards that this was even possible. Meeting these standards have not been easy, however, and I owe a debt of gratitude to so many people who have supported me in so many ways—each member of my esteemed doctoral committee who were generous with their time and expertise, former and current Odum Institute colleagues who allowed and encouraged me to take on this pursuit, friends who gave me reassurances right when I needed it, and my family who have been accepting and appreciative of my lifetime student status. To Helen Tibbo, thank you for guiding me through my research program (while also giving me the best life advice). To John D. Martin III, thank you for being by my side during the hardest part of this journey—I could not have done this without you. To Erin Henderson, thank you for making me feel as smart as you are. To Peter Leousis, thank you for believing in my potential enough to invest in me. To the late Thomas Carsey, for opening the door to this area of research. To Jonathan Crabtree, thank you for entertaining my big ideas. To so many other friends, family, and colleagues who inspire me every day, thank you. Finally, to Ben August, thank you for being the reason that I do what I do.
# TABLE OF CONTENTS

LIST OF TABLES .................................................................................................................. x

LIST OF FIGURES ............................................................................................................... xi

CHAPTER 1: INTRODUCTION ................................................................................................. 1

  1.1 Journal-based Data Policy ............................................................................................ 2

  1.2 Problem Statement ....................................................................................................... 5

  1.3 Dissertation Overview ................................................................................................ 6

CHAPTER 2: LITERATURE REVIEW ....................................................................................... 8

  2.1 The Reproducible Research Imperative ...................................................................... 8

  2.2 The Reproducibility Standard ...................................................................................... 13

  2.3 The Reproducible Research Mandate ......................................................................... 15

     2.3.1 Community Mandates ......................................................................................... 16

     2.3.2 Funder Mandates ............................................................................................... 18

     2.3.3 Journal Mandates ............................................................................................. 20

  2.4 The State of Reproducible Research Practices .......................................................... 23

  2.5 Theoretical Framework ............................................................................................... 27

     2.5.1 Theory of Planned Behavior ................................................................................. 27

     2.5.2 Theory of Planned Behavior and Reproducible Research Practices ................. 30

  2.6 Literature Review Conclusion ..................................................................................... 31
CHAPTER 3: RESEARCH METHODS .............................................................................................................. 33

3.1 Study Design ........................................................................................................................................ 33

3.1.1 Constructivist Grounded Theory ................................................................................................. 34

3.1.2 Constructivist Grounded Theory and TPB .................................................................................. 38

3.2 Document Content Analysis ................................................................................................................ 39

3.2.2 Document Sampling ........................................................................................................................ 41

3.2.3 Document Coding ........................................................................................................................... 42

3.2.4 Statistical Analysis ............................................................................................................................ 49

3.3 Qualitative Interviews .......................................................................................................................... 50

3.3.1 Participant Sampling ....................................................................................................................... 51

3.3.2 Qualitative Data Collection ............................................................................................................. 53

3.3.3 Qualitative Data Analysis ................................................................................................................ 54

3.4 Study Validity ......................................................................................................................................... 57

3.5 Role Considerations ............................................................................................................................. 58

CHAPTER 4: RESEARCH RESULTS ............................................................................................................. 59

4.1 Content Analysis Results ..................................................................................................................... 59

4.2 Qualitative Interview Analysis Results ............................................................................................... 61

4.2.1 Beliefs About Reproducible Research ............................................................................................ 61

4.2.2 Reproducible Research Behaviors ................................................................................................. 67

CHAPTER 5: DISCUSSION AND CONCLUSION ......................................................................................... 73
# LIST OF TABLES

Table 1. Manuscripts by resubmissions .........................................................................................42
Table 2. Access error codes .............................................................................................................43
Table 3. Documentation error codes ...............................................................................................45
Table 4. Computation error codes ..................................................................................................48
Table 5. Interview sample characteristics ......................................................................................52
Table 6. Manuscripts by error type ..................................................................................................59
Table 7. Results of OLS regression ...............................................................................................60
Table 8. Behavioral beliefs affecting reproducible research practices ...........................................62
Table 9. Normative beliefs affecting reproducible research practices ...........................................64
Table 10. Control beliefs affecting reproducible research practices ..............................................66
Table 11. Actual behavioral control factors affecting reproducible research practices ..................67
LIST OF FIGURES

Figure 1. AJPS data policy implementation workflow ................................................. 3
Figure 2. Distinction among terms referring to benchmarks of scientific integrity .......... 9
Figure 3. Theory of Reasoned Action model (Azjen, 1985) ............................................. 28
Figure 4. Theory of Planned Behavior model (Azjen, 2019) ........................................... 30
Figure 5. Theory of Planned Behavior applied to data sharing behavior ......................... 31
Figure 6. Components of actual behavioral control in the context of journal-based data policy compliance ................................................................................................. 74
CHAPTER 1: INTRODUCTION

Despite general consensus within the research community that reproducibility is imperative to scientific integrity and progress, its members have not been engaging in the necessary activities to ensure that their own published research is reproducible (Andreoli-Versbach & Mueller-Langer, 2014). The research artifacts, i.e., the data, code, documentation, and other materials that serve as the evidence-base for published scientific findings, frequently remain unavailable despite the many repositories that offer convenient web-accessible platforms for sharing and publishing these artifacts (Peer et al., 2014). Even repositories cannot make guarantees as to the quality of artifacts placed under their careful stewardship. Inadequate documentation, undefined dataset variables, and code riddled with scripting problems all have been found in repositories (Stodden et al., 2018a).

Reproducibility relies on scientists to share the research artifacts that support their published claims while also taking care that other scientists can make use of those artifacts to repeat the computational steps of their analysis and generate the same results (National Academies of Sciences, Engineering, and Medicine, 2019). Collectively, these files comprise the research compendium (Gentleman & Lang, 2004) that should be “independently understandable for informed reuse” (Peer et al., 2014). However, the time and effort needed to assemble and describe requisite files in the compendium, document analysis steps, clean datasets, test analysis code, and publicly share these materials is not something many scientists are willing to spare without the promise of personal benefits (Kim, 2010; Tenopir et al., 2011,
This being the case, research stakeholders have created incentives to compel scientists to perform these reproducible research practices to allow for verification of research results. For academic researchers whose career advancement is often tied to publication, journal-based policies that require authors to submit research artifacts underlying reported findings that are evaluated for quality as a condition of publication ought to be persuasive (Crosas et al., 2018; Elman et al., 2018; Fecher, Friesike, & Hebing, 2015; Key, 2016; Santori, 2016).

1.1 Journal-based Data Policy

In 2015, the *American Journal of Political Science* (AJPS), issued a data verification policy stipulating that manuscript publication will be withheld until the author has submitted a research compendium that reproduces the analytical results reported in the manuscript (American Journal of Political Science, n.d.; Jacoby et al., 2017). The language of the AJPS policy made it clear to authors that they “must provide materials that are sufficient to enable interested researchers to verify all of the analytic results that are supported in the text and supporting materials,” and that “the materials will be verified to confirm that they do, in fact, reproduce the analytics results reported in the article” (American Journal of Political Science, n.d.). Even though the policy is unambiguous as to its requirements, and even with additional guidance documents available to authors to support their efforts to meet policy requirements (American Journal of Political Science, 2015), the vast majority of authors’ efforts are nevertheless unsatisfactory based on the documented results of independent verification performed by data curation and statistical computing experts (Christian et al., 2018).

The AJPS policy has been successful in that the research compendia for all articles published in the journal since the policy took effect can be accessed from a public repository housed in the Harvard Dataverse (apart from a handful of articles for which the author was
granted a policy exception due to the sensitive or proprietary nature of their analysis data).

Implementation of the policy has not been without its difficulties, particularly those from policy non-compliance. If an author submits their compendium and independent verifiers discover deficiencies in the compendium, the author is then prompted to make prescribed corrections to their materials and submit them for subsequent re-verification. Figure 1 presents a high-level illustration of the verification workflow.

Figure 1. AJPS data policy implementation workflow.

Despite specific policy guidelines for preparing specified files and their contents, it is common for authors to submit (and resubmit) compendia that have missing or incomplete
materials precluding any ability to verify reproducibility, analysis code that fails to execute due to its non-portability to a third-party computing environment, and/or inadequate documentation making the analysis workflow indeterminable. These correctable but disqualifying issues can extend the compendium submission-verification-resubmission-reverification process considerably, making feasibility of implementing such a strict policy doubtful, particularly for those journals operating in tightly resourced and/or time-sensitive environments (Crotty, 2016).

For AJPS, the introduction of their data verification policy extended the time to publication by a non-trivial 53 days, with the number of hours required for successful verification adding up to approximately six to eight hours on average for each manuscript (Jacoby, 2015; Jacoby et al., 2017). These time and cost expenditures are significant, pressuring AJPS editors and their collaborating verifiers to analyze the verification workflow in search for ways to streamline the process for maximum efficiency. The current process involves labor-intensive inspection of every compendium file, every variable in the dataset, and every line in the code. It also includes execution of every code command and comparison of every code output with every result in every line of text, table cell, and/or graphic image (Christian et al., 2018).

While the workflow analysis has pinpointed some activities that can be streamlined through standardization or automation, the single factor that drives up verification time, and therefore policy implementation costs, is the repetition of the compendium submission and verification workflow instigated by authors’ failure to meet the policy requirements for reproducible research compendium submission.

Implementation of the AJPS data policy takes place at the Odum Institute at the University of North Carolina at Chapel Hill, which performs third-party verification for AJPS manuscripts accepted for publication pending successful verification. As the assistant director of archives at the Odum Institute, I have overseen the execution of data curation and verification
processes for over 500 manuscripts. Since beginning this work with AJPS when the policy took effect, the Odum Institute verification team has looked for ways to increase authors’ compliance with policy requirements. Refining policy language to further specify compendium requirements, developing guidance documents to support compendium preparation, and offering detailed suggestions for correcting compendium deficiencies have gone only so far to reduce the amount of time and labor spent ensuring that the research materials supporting published analytical results meet the reproducibility standard. Even now, manuscripts still require about two resubmissions on average to successfully pass verification.

1.2 Problem Statement

To understand why it is the case that authors do not submit reproducible research compendia even when policy requires it (and policy guidance explains how), it is worth looking toward the specific mechanisms that inhibit data management activities that support reproducibility.

Previous studies on data sharing behavior, which is requisite for reproducibility, have pointed to specific determining factors for data sharing that align with the well-established theory of planned behavior (Ajzen, 1985; Ajzen, 2011). In general, according to those studies, given personal incentives (attitude toward the behavior) social expectations and/or regulatory pressure (subjective norms), and ability (perceived behavioral control), scientists will share their research data (Ajzen, 2019). It should follow, then, that if a scientist had an article published in a journal (incentive) that had a data sharing policy in place when they submitted their article for review (regulatory pressure) and was capable of sharing their data (perceived behavioral control), then the data should be publicly accessible. Yet, other studies that have sought to investigate the effectiveness of these policies have found this to not be always the case.
Scientists’ failure to perform the desired data management behaviors despite the presence of factors known to predict behavior likewise describes the experience with the AJPS policy implementation.

Further investigation into the mechanisms that encourage and/or inhibit data management behaviors can inform the design of interventions that better equip researchers to produce and share high-quality research compendia, thus reducing the amount of time and labor currently required for current data verification workflows. This, in turn, may make journal-based data policies more feasible for a greater number of journals.

To be sure, there has been a small number of authors whose research compendia did not require resubmission of modified materials. Whether their compliance had to do with their clearer understanding of the terms of the policy, approval of the policy’s aims, ability to fulfill policy requirements, avoidance of the consequences of non-compliance, or any other number of factors is yet unknown. In any case, researchers who have been subject to journal-based data policies that require reproducible research compendium sharing as a condition of publication—that is, researchers who are known to have executed the tasks required to share high-quality data for the purpose of allowing independent verification of the reproducibility of their reported research results—are well-positioned to offer more precise insights into the factors that affect data management and sharing behavior and how journal-based policies work to promote these behaviors.

1.3 Dissertation Overview

This dissertation describes a mixed-methods grounded theory (MM-GTM) research study that considers the implementation of the AJPS data policy as an exemplar of a rigorous journal-based data policy that serves as an apparatus for compelling scientists to adopt data management
practices that support research reproducibility. Corresponding with previous studies, the
dissertation applying the theory of planned behavior as sensitizing concepts to confirm, clarify,
and discover factors that explain data behaviors in the context of journal-based policies.
Specifically, the study addresses the following research questions:

- What are the common causes of non-reproducibility of research compendia?
- Why do researchers fail to produce reproducible research compendia?
- How does the journal-based data policy work to promote reproducible research
  practices?

Among the presentation of study findings, the dissertation proposes an expansion to the
theory of planned behavior framework to explain how the policy operates to encourage adoption
of reproducible research practices. Finally, the dissertation discusses how the study findings and
the proposed framework may be used to inform interventions that support policy compliance and
adoption of reproducible research practices.
CHAPTER 2: LITERATURE REVIEW

This literature review covers scholarly works that form the knowledge base for understanding the phenomenon of this investigation as well as exposing gaps in knowledge that motivated the dissertation research. The literature review begins by illustrating the persistence of verifiability throughout history as a central tenet of scientific practice, and the operationalization of verifiability as reproducibility when applied to research approaches that rely on the computation of quantitative data for scientific discovery. This is followed by discussions about the reproducibility standard and the challenges of upholding this standard despite key stakeholder efforts to enforce it through policy mandates. Finally, I outline what is known about the current state of reproducible research practices and the theoretical framework that informed the research methods used in the study.

2.1 The Reproducible Research Imperative

While the topic of data sharing and research reproducibility has gained prominence in recent years, the expectation that products of research be made available to the scientific community to allow for verification of claims was established long before journals issued data verification policies. Verifiability is known to be a central tenet of the modern scientific method that appeared as early as the eleventh century in the writings of Ibn al-Haytham in which he described the necessity of systematized experimentation, testing, and independent verification (Bettany, 1995; Tbakhi & Amr, 2007). The scientific principle of verifiability has been a constant throughout history, also making its appearance as the fourth step of Roger Bacon’s cycle of scientific research after observation, hypothesis, and experimentation, which is
considered the closest precursor to the modern scientific method (Nosek et al., 2012). In more recent history, Robert Merton (1973) introduced the term *organized skepticism* to describe verification as one of four imperatives for sound scientific practice, known today as the Mertonian Norms. The National Academies (2019) recently published the report, “Reproducibility and Replicability in Science,” which associated scientific rigor with verifiability. Even the general public is attuned to this in some sense, with over half of Americans indicating in a survey that they are more likely to trust science if researchers make their data openly available and are subject to independent verification (Funk et al., 2019).

It should be noted that reproducibility is but one benchmark of scientific integrity when considering computational research. The practice of verifying scientific claims can take on various forms, which often are based on disciplinary norms and standards. Scholars have made distinctions among terms such as *robustness, replicability, generalizability, and reproducibility* (among others) to explain the importance of each to scientific rigor. The Turing Way (2019) used these terms to distill the assortment of signifiers of existing benchmarks of scientific integrity distinguished by noting the use of the same or different data and analysis as the original study for validation (see Figure 2).

<table>
<thead>
<tr>
<th>DATA</th>
<th>same</th>
<th>different</th>
</tr>
</thead>
<tbody>
<tr>
<td>same</td>
<td>REPRODUCIBLE</td>
<td>REPLICABLE</td>
</tr>
<tr>
<td>different</td>
<td>ROBUST</td>
<td>GENERALIZABLE</td>
</tr>
</tbody>
</table>

**Figure 2.** Distinctions among terms referring to benchmarks of scientific integrity (The Turing Way Community et al., 2019).
Each of these terms refer to a practice that seeks to determine the veracity of research results using a systematic approach for critically assessing research protocols and analyses while redressing potential errors (Plesser, 2018).

For empirical research that relies on computational approaches to produce research results, verifiability is often associated with *reproducibility*, which presumes that, given the same data and code used in the original computation, analysis steps can be repeated to yield the same results (National Academies of Sciences, Engineering, and Medicine, 2019; Stodden, 2013, 2015). In a 2007 lecture, Jim Gray declared the arrival of a *fourth paradigm* of scientific discovery characterized by the emergence of computational science made possible by new technologies for capturing and analyzing massive amounts of complex data. Whereas bench scientists record their experimental protocols and observations in lab notebooks, much of the data processing and analysis workflows executed in computational science are captured in programming scripts (Hey et al., 2009). As such, these captured workflows can be audited, debugged, and re-executed (Goble & De Roure, 2009). This should make it more likely that research findings that rely on computation can be reproduced.

Yet, questions have been raised about the credibility of science in response to reports of failed attempts to reproduce computational findings reported in frequently-cited research papers. One noteworthy example is the Reinhart-Rogoff controversy (Cassidy, 2013), which spotlighted serious problems with research findings that stood for years as the basis for public policy in countries around the world. In their applied econometrics graduate course at the University of Massachusetts, Amherst, professors Michael Ash and Robert Pollin assigned a term paper in which students were to present the outcome of their attempts to reproduce findings reported in a published empirical econometrics article. When Thomas Herndon selected the seminal Reinhart and Rogoff (2010) paper, “Growth in a Time of Debt,” neither he nor his professors anticipated
the discovery of significant errors in the original authors’ computation of the relationship between public debt and GDP growth. Reinhart and Rogoff’s flawed results, which claimed that increases in government debt at a certain point are associated with substantial decreases in economic growth, were referenced by prominent policymakers to support arguments for austerity measures (as opposed to fiscal stimulus) in countries struck by recession.

Unfortunately, the Reinhart-Rogoff case is not unique. In recent years, several investigators have found that a significant number of published findings they examined cannot be reproduced, and many of those that can be reproduced require a considerable amount of effort. In many of these studies, researchers were unable to access the data needed to verify reproducibility in the first place (Bergh et al., 2017; Giofrè et al., 2017; Vanpaemel et al., 2015; Vlaeminck & Herrman, 2015; Zenk-Möltgen & Leptien, 2014). Bergh, Sharp, Aguinis, and Li (2017) were able to reproduce full or partial results for just 26 out of 88 articles published in the Strategic Management Journal, with the remainder unable to be tested due to lack of data availability. In Stodden, Krafczyk, & Bhaskar’s (2018b) tests for computational reproducibility on articles for which they were able to obtain the data, code, and other research artifacts necessary to regenerate results, they discovered that only about half were reproducible, with none able to be reproduced without some notable degree of effort and expertise. Konkol, Kray, & Pfeiffer (2018) encountered similar results in their study in which findings reported a mere two out of 14 papers they inspected were reproducible.

According to Peng (2015), these examples of reproducibility failure suggest a scientific enterprise in crisis that warrants a review of the scientific method and its ability to uphold the integrity of the scientific record. It should be noted, however, that only in rare cases is irreproducibility the result of scientific misconduct by way of data fabrication, falsifying results, or other flagrant misdeeds (Eisner, 2018; Fanelli, 2018; Franzen, 2016). Computational research
is prone to error (Donoho, 2010; Donoho et al., 2009), with reproduction failures more often the result of scientific practices that neglect the data management tasks necessary to enable others to retrace analytical steps for generating outputs corresponding to reported results.

Inadequate documentation resulting in analytical missteps, inadvertent use of incorrect versions of dataset files, unspecified computing environment parameters leading to output discrepancies from use of dissimilar technologies, human errors in variable coding, irregular file or variable names causing code execution errors, absence of non-executable comments that render code blocks inexplicable, failure to transfer research materials to a repository for public sharing—all of these are examples of errors that create impediments to computational reproducibility (Brown et al., 2018; Buckheit & Donoho, 1995; Donoho et al., 2009).

Regardless of the reasons for irreproducibility, the Reinhart-Rogoff and similar cases have prompted calls for data access and research transparency, which are prerequisites for reproducibility, to be integral to normative research practice (Borgman, 2010, 2012; Darch & Knox, 2017; Stodden, 2015; Tenopir et al., 2011). By relying on the published article to verify scientific claims, stated Franzen (2016), “...we can only look at the front-stage component of research, the representational side. The backstage, i.e., the knowledge production component, however, remains in the dark” (p. 473). Goodman et al. (2016) referred to this as “blind alleys" (p. 5) in which researchers find themselves where information to enable evaluation of reported findings is missing or insufficient. As Lupia and Elman (2014) noted, transparency allows audiences to more fully understand the approaches and contexts in which research findings are produced so that they can form an evidentiary and logical basis for treating the claims as valid” (p. 22). Otherwise, Munafò et al. (2017) explained, the credibility of the findings is dependent solely on trust in the researcher reporting the finding, but that “[t]ransparency is superior to trust” (p. 5).
2.2 The Reproducibility Standard

In response to the concerns raised with respect to highly publicized cases of failure to reproduce, individuals and groups of researchers have made various calls to action in support of the restoration of scientific credibility. Among these calls have been an insistence to instill the principles of research reproducibility into community culture, adopt transparent research practices in support of reproducibility, and hold peers accountable for adopting these practices while offering incentives to those who do.

Reproducibility of reported research results is predicated on the availability of what Gentleman & Lang (2004) first referred to as a “compendium” of research artifacts crucial for reproduction. The research compendium fills in essential details about the backstage knowledge production activities that tend to be missing from the methods section of the scientific article (Franzen, 2016). About scientific publication, Claerbout is said to have called it “advertising of the scholarship,” as opposed to the “actual scholarship,” which exists as the research compendium (Buckheit & Donoho, 1995). Claerbout & Karrenbach (1992) wrote that “the burdens imposed on the author to create reproducible results are little more than the task of filing everything systematically” (p. 604). Indeed, the research compendium bundles the data, code, documentation, and all other research artifacts necessary to generate analytical results.

Beyond assembling requisite files, however, data management tasks also must be performed so that the materials in the compendium meet minimum standards of quality that enable an individual apart from the researcher(s) who performed the original study to retrace the analytical steps to achieve identical results (Gentleman & Lang, 2004; Piccolo & Frampton, 2016). In his seminal article, “Replication, Replication,” Gary King (1995) defined a
reproduction\(^1\) standard, which “holds that sufficient information exists with which to understand, evaluate, and build upon a prior work if a third party could replicate the results without any additional information from the author” (p. 444). To translate this standard to concrete tasks, Peer, Green & Stephenson (2014) proposed a framework for data quality review that includes a number of data management activities for ensuring the completeness and quality of files, documentation, data, and code for the purpose of supporting reproducibility.

Unfortunately for many researchers, data management is not well-known or understood (Ward et al., 2011). Core graduate school curricula rarely include comprehensive data management training, leaving researchers without essential skills for producing and sharing high-quality research compendia (Frugoli et al., 2010; Larigot, 2019). Researchers, both early career and well-established, end up relying on insufficient ad hoc strategies to deal hastily with data management challenges (Frugoli et al., 2010). For many of these researchers, the prospect of adding tasks to their already burdensome workloads to prepare and assemble research research compendia for sharing, albeit vital to the integrity of their research, is a daunting one given the amount of time and effort they perceive these tasks to require (Fecher, Friesike, & Hebing, 2015; Gherghina & Katsanidou, 2013; Tenopir et al., 2011, 2015).

Moreover, investigations that involve human subjects research are subject to laws and regulations that place restrictions on data dissemination, which further complicates efforts to share data. To avoid sanctions from unknowingly failing to comply with sensitive data handling protocols to safeguard the privacy of study participants, researchers will instead opt to destroy the data upon completion of their analyses to avoid risk of unintended disclosure of sensitive information, rather than take measures to de-identify the data and assume unavoidable disclosure.

---

\(^1\) Here, King uses the term “replication” to refer to computational reproducibility, which was the common usage in the political science domain in which he practices research.
risks (Akers & Doty, 2013; Alnoamany & Borghi, 2018; Kim & Stanton, 2015). Similarly, investigations that make use of proprietary data from commercial sources are often subject to license restrictions or intellectual property laws that disallow redistribution of the data and/or the software required to make use of them (Borgman, 2010; Elman et al., 2018; Faniel & Yakel, 2017; Freese & Peterson, 2017; King, 2011). Stodden et al. (2016) reiterated these challenges, conceding that some ingenuity would be required to create workarounds that likely would involve infrastructure, policies, and procedures to provide access to restricted materials.

Should researchers expend the effort to share their data (and to perform the data management tasks and/or workarounds that make data sharing possible), they will likely find that their efforts go largely unrewarded. Despite a general consensus that data sharing is essential to scientific advancement, there is very little, if any, personal incentive for researchers to engage in the labor-intensive process of packaging and distributing their data (Arzberger et al., 2004; Borgman, 2010, 2012; Lupia & Elman, 2014; Nosek et al., 2015; Pienta et al., 2010; Wallis et al., 2013). In other words, researchers do see the desirability and profitability of data sharing for their field, but less so for their own research projects (Houtkoop et al., 2018).

2.3 The Reproducible Research Mandate

Regardless of the challenges researchers face in upholding the reproducibility standard, key stakeholders of the scientific enterprise have taken note of reproducibility failures and their consequences. In response, they have declared reproducibility an endeavor worthy of the additional time and effort it requires of researchers. To compel researchers to perform the activities that support research reproducibility, professional academic societies have updated their guides to professional ethics to include language that articulates community expectations for transparent research practices. Funding agencies from both public and private sectors have
thus made it their policy that grant recipients share the research artifacts generated during the
course of sponsored research activities. Scholarly journals, which researchers rely on to
disseminate and garner recognition for their research outputs have issued policies requiring
authors to share the research compendia used to generate reported results as a condition of
publication.

2.3.1 Community Mandates

The insistence on making research transparent has been a mainstay within research
communities, as evidenced in the codes of ethics of academic societies that work to promote the
activities of their respective disciplinary communities (Delicado et al., 2014). The Royal
Society, which was founded in 1660 and known as the oldest academic society in the world,
underscored the indispensability of evidence to scientific knowledge-making with its motto,
“Nullius in verba,” or “take nobody’s word for it” (The Royal Society, n.d., para. 2; Srivastava,
2018). This ethos is reflected in the summary of the Royal Society’s 2012 report, “Science as an
Open Enterprise”:

Open inquiry is at the heart of the scientific enterprise. Publication of scientific
theories--and of the experimental and observational data on which they are based--
permits others to identify errors, to support, reject or refine theories and to reuse data for
further understanding and knowledge. Science’s powerful capacity for self-correction
comes from this openness to scrutiny and challenge (Royal Society & Policy Studies
Unit, 2012, p. 7).

Similar statements regarding open data and research transparency appear in documents
issued by other contemporary academic societies across disciplinary domains. In 2012, the
American Political Science Association’s (APSA) revised its Guide to Professional Ethics in
Political Science to include a series of new statements making data access and research transparency an ethical standard of the profession: “Researchers have an ethical obligation to facilitate the evaluation of their evidence-based knowledge claims through data access, production transparency, and analytic transparency so that their work can be tested or replicated” (APSA Committee on Professional Ethics, Rights and Freedoms, 2012, p. 9). According to Lupia and Elman (2014), what prompted these revisions was growing concern from the APSA’s governing council over instances of data withholding and non-reproducible published findings. After several discussions with APSA committees and members, new language making clear the political scientist’s obligation to engage in transparent research practices was added to the Guide to Professional Ethics in Political Science.

Likewise, the American Geophysical Union (2015) issued a position statement with the declarative title, “Earth and Space Science Data Should be Credited, Preserved, Open, and Accessible as an Integral Responsibility of Scientists, Data Stewards, and Sponsoring Institutions.” The American Economic Society (AEA) adopted the AEA Code of Professional Conduct, which begins with a statement that refers to “intellectual and professional integrity” (American Economic Association, 2018, para. 2) that requires, among other things, research transparency. Included in the American Psychological Association’s (2016) Ethical Principles of Psychologists and Code of Conduct is a statement disavowing the practice of withholding data associated with reported findings from researchers seeking to verify those findings. In the General Ethical Principles of the ACM Code of Ethics and Professional Conduct (2018), the Association for Computing Machinery (ACM) advises its members to support other computing professionals by making their work open and freely available to others as a public resource.
2.3.2 Funder Mandates

In any discussion of data sharing policies, it is rare not to see a mention of the 2010 National Science Foundation (NSF) press release announcing revisions to its proposal submission policy. This press release explained the new data management plan component of the proposal package that required researchers to outline a strategy for managing and disseminating data generated in the course of the proposed research project. The reason for the new requirement was rationalized in part by Ed Seidel, then acting director for the NSF Mathematical and Physical Sciences directorate. Quoted in the press release, he explained, “Science is becoming data-intensive and collaborative. Researchers from numerous disciplines need to work together to attack complex problems; openly sharing data will pave the way for researchers to communicate and collaborate more effectively” (National Science Foundation, 2010, p. 1).

In 2013, the Office of Science and Technology Policy (OSTP) made it clear in a memorandum to all government agencies their expectation that outputs from federally funded research be disseminated to the public “to maximize the impact and accountability of the Federal research investment” (Holdren, 2013), among other reasons. Accordingly, several government agencies including the Centers for Disease Control (2016), NASA (2014), the Department of Energy (2014), the Department of Transportation (2015), and the Department of Defense (2018) issued new data policies or updated existing data policies not long after the release of the OSTP memorandum.

More recently, the National Institutes of Health (NIH) made important updates to their 2003 Data Sharing Policy to not only require investigators to include in their grant proposals a data management and sharing plan that includes provisions to share data upon publication of studies that use the data, but also to demonstrate that the data were shared according to the plan.
Reissued as the Data Management and Sharing Policy to take effect January 2023, NIH cited the reason for the new policy, which is “...to promote the sharing of scientific data. Sharing scientific data accelerates biomedical research discovery, in part, by enabling validation of research results, providing accessibility to high-value datasets, and promoting data reuse for future research studies” (National Institutes of Health, 2020).

With the 2022 release of another OSTP memo addressing the need to make publicly funded research assets public, other government agencies will be required to follow NIH’s lead and update their own policies to mandate the timely release of data resulting from federally funded research (Nelson, 2022). This memo, titled “Ensuring Free, Immediate, and Equitable Access to Federally Funded Research,” stated the reason for this requirement:

Improving public access policies across the U.S. government to promote the rapid sharing of federally funded research data with appropriate protections and accountability measures will allow for greater validity of research results and more equitable access to data resources aligned with these ideals (Nelson, 2022, p. 2).

This statement goes one step further than Holdren’s 2013 OSTP memo, which places emphasis on maximizing federal investments in science, toward focused attention on the integrity of scientific research.

Researchers whose projects are supported by private funders have not been immune to such mandates. While government agencies have been working to develop policies to comply with the OSTP directives, several private-sector funders also have been requiring benefactors of their research investments to engage in data management and sharing activities for the same reasons given by their government counterparts. Private philanthropic organizations with smaller research investment portfolios such as the Andrew W. Mellon Foundation (2011),
Gordon and Betty Moore Foundation (2014), Gates Foundation (2015), Ford Foundation (2015), MacArthur Foundation (2015), and Alfred P. Sloan Foundation (2021) all have issued their own policies that make it a requirement that research artifacts be shared.

2.3.3 Journal Mandates

Journal policies are of particular interest to the dissertation study due to their potential to effect change in normative research practices by compelling researchers to adopt data management and sharing activities that support research reproducibility as a condition of publication. Of the policies issued by key scientific stakeholders, journal-based policies have the greatest impact on data sharing behavior, according to studies by Enke et al. (2012), Kim & Stanton (2015), and Houtkoop et al. (2018). These findings support claims that journal editors are in a unique position within the typical academic reward structure that privileges article and book publication over other metrics of research productivity (Zenk-Möltgen et al., 2018).

Because academia places significant weight on scholarly publication in tenure and promotion decisions, researchers are at the mercy of journal guidelines for manuscript submission and acceptance (Elman et al., 2018; King, 1995; Munafò et al., 2017).

Editors with a discerning eye for robust, reliable research act as gatekeepers responsible for protecting the integrity of the journal (Franzen, 2016; McCain, 1995; Nosek et al., 2012; Zenk-Möltgen et al., 2018). McCain (1995) referred to their role as one of gatekeeper in which editors establish standards of quality and relevance that determine what is worthy of publication. Policies issued by editors that make publication contingent on data policies make researchers accountable for their scientific outputs by holding them to a standard of transparency and openness (Nosek et al., 2012), thus predisposing them to sharing their research compendia (Fecher, Friesike, & Hebing, 2015; Zenk-Möltgen et al., 2018). Use of “prepublication controls” (McCain, 1995, p. 416) such as data sharing policies extends the journal’s gatekeeping role to
compel researchers to adopt research practices seen by editors to maintain and reinforce their journals’ standards. In this position, editors have authoritative power to issue policies that make publication contingent on authors sharing their data publicly to allow for confirmation of reproducibility of published results.

Considering the role the journal can play in promoting research reproducibility, the Transparency and Openness Promotion (TOP) committee, comprised of leading scholars, journal editors, and funding agencies, developed a common set of standards “to translate scientific norms and values into concrete actions and change the current incentive structures to drive researchers’ behavior toward more openness” (Nosek et al., 2015, p. 1423). These resulting TOP guidelines define a set of standards for openness that journals can adopt as policy and at various levels of stringency. In a list of signatories last updated in 2020, over 5,000 journals declared their support for the TOP guidelines and committed to reviewing the alignment of their policies with the guidelines (Center for Open Science, 2015).

Despite the expectation that journal-based data policies will compel researchers to engage in data sharing behaviors, it is a common conclusion in the literature that this is not always the case. Wicherts, Borsboom, Kats, and Molenaar (2006) contacted the authors of 141 articles published in American Psychological Association (APA) journals--assuming the authors had signed the APA Certification of Compliance with Ethical Principles, which includes language making it obligatory for researchers to share their data. In response to 400 emails, they received datasets for only 64 studies (26%). Savage and Vickers (2009) requested data from authors who published in PLOS Medicine or PLOS Clinical Trials, both of which had an explicit data sharing policy in place. Of the ten authors to whom they sent email requests, only one shared their data on the first request. Four others declined to share and three did not respond, while two email addresses were found to be no longer valid. While their sample was small, Savage and Vickers
still maintained their results to be indicative of the ineffectiveness of data sharing policies. For 351 articles published in high-impact journals with a data availability policy, Alsheikh-Ali, Qureshi, Al-Mallah, and Ioannidis (2011) found 143 (41%) with analysis data publicly accessible. When Stodden, Seiler, and Ma (2018b) sought to obtain data and code for 204 randomly selected articles in Science, they were able to procure the materials for 24 articles by using access information provided in the article. They obtained data and code for an additional 65 articles by requesting the materials from the authors for a success rate of 44%.

These studies did not ascertain whether the results would have been the same if a policy were not in place. Consider the study by Stockemer, Koehler, and Lentz (2018), in which the investigators were able to access data for 57% of articles published in journals without a data policy. However, a study by Vines et al. (2013) that compared outcomes between journals with data policies and journals without data policies found that journals with a policy that required authors to submit their data to a repository resulted in a higher probability that the data were available online. Also addressing this uncertainty, Hardwicke et al. (2018) conducted a pre- and post-policy adoption experiment on articles in the journal Cognition to discover whether there was an increase in the presence of data availability statements and the degree to which data were apparently usable. Indeed, the investigators found an increase in data availability statements from 25% pre-policy to 78% post-policy with 22% of the pre-policy data appearing usable compared to 62% of post-policy data.

The reasons for non-compliance with journal-based data policies certainly may vary, but several scholars have noted that the effectiveness of these policies depend on their stringency, with policies using stronger language clearly indicating sharing to be a “requirement” rather than a “recommendation” or “expectation” yielding higher rates of compliance (Christian et al., 2020; Piwowar & Chapman, 2010; Vines et al., 2013; Vlaeminck & Herrman, 2015). The most
effective policies also include mechanisms for enforcement because, as Crotty (2016) wrote, “A policy without teeth--without actual consequences for non-compliance--loses all effectiveness” (p. 2).

But even for some journals that have implemented the most stringent of data policies, such as that of AJPS, compliance is still uneven, with the overwhelming majority of researchers initially failing to meet specified policy requirements designed to ensure that shared materials can be used to reproduce published findings (Janz, 2015). A closer look at researchers’ reproducible research practices can shed some light on why this is the case.

2.4 The State of Reproducible Research Practices

To determine the degree to which researchers have been engaged in the reproducible research practices, several scholars have carried out empirical studies of researchers in a variety of disciplinary domains. Many of these studies focus on data sharing, given that the scientific record cannot assert truth without first providing access to the evidence underlying published research findings (Stark, 2018). According to studies that appraised the current state of data sharing, which is a principal activity for supporting reproducibility, and investigated the factors that motivate researchers to engage in this reproducible research practice, data sharing has yet to be established as a normative research practice within and across many disciplines (Andreoli-Versbach & Mueller-Langer, 2014; Fecher, Friesike, & Hebing, 2015; Fecher, Friesike, Hebing, et al., 2015; Houtkoop et al., 2018; Huang et al., 2012; Jeng et al., 2016; Zenk-Möltgen et al., 2018).

And still, according to these same and other studies (Schmidt et al., 2016; Tenopir et al., 2011, 2015), scientists do tend to agree that data sharing is essential to the research enterprise. The incongruity between attitudes toward data sharing and participation in data sharing activities
has prompted questions as to why researchers agree that data sharing is important to science, while at the same time eschewing the practice for themselves. Several scholars have sought to answer these questions by examining the characteristics of individual researchers as well as the environment in which they share (or do not) their data.

To investigate the determinants of data sharing behavior, some scholars, albeit few, have applied theoretical frameworks to help address the dissonance between individuals’ acknowledgement of the importance of behavior and their actual performance of the behavior. Jeng, He, and Oh (2016) applied to their investigation the knowledge infrastructures (KI) conceptual framework and the theory of remote scientific collaboration (TORSC). As they explain, KI describes a complex ecosystem of people, norms, artifacts, institutions, practices, policies, and technology. TORSC, which defines concepts for successful scientific collaboration (common ground, coupling work, collaborative readiness, management, and technology readiness) is applied to data sharing, which the authors consider to be a type of scientific collaboration. Jeng, He, and Oh combine and distill these two theoretical frameworks into two primary dimensions that influence data sharing behavior: individual (individual motivations and characteristics), and contextual (data characteristics, organizational and research culture, and technical infrastructure). Results reiterated the majority positive attitude toward data sharing in contrast to a lack of actual data sharing behavior. Because of this, the authors suggested the need for further study on the specific barriers that impede or disallow data sharing.

Pronk, Wiersma, van Weerden, and Schieving (2015) operationalized data sharing factors into costs and benefits as part of a game theoretical problem for determining “strategic choice” (p. 2) to share data. Presuming readers’ prior understanding of game theory, the authors do not provide an explicit definition except to allude to the game theory concept of “the tragedy of the commons” (p. 2) as analogous to data sharing. That is, unless the scientific community as a
whole agrees to participate in data sharing, the cost to an individual researcher to share their data outweighs the benefits, rendering the decision to share disadvantageous. Adjusting for time-cost and citation advantage parameters in a game theoretical mathematical expression, the authors were able to calculate, using simulations, how those parameters are likely to affect data sharing decisions. Based on their calculations, the authors found policies to play an essential role in promoting data sharing, with a decrease in time and effort and an increase in incentives likely to have a positive effect on sharing.

The theory of planned behavior (TPB) appears prominently in several studies seeking to determine the factors that affect data sharing behavior. Kim and Zhang (2015) applied constructs from TPB to their survey of scientists from various disciplines in an effort to identify the determinants of data sharing behavior. In this application, the authors operationalized attitude toward the behavior as being based on scientists’ perceived career benefit and risk from sharing their data, and perceived effort required to share their data. Subjective norms were based on individuals’ understanding of their colleagues’ expectations for them to share their data. Perceived behavioral control was defined in this study as the availability of data repositories to facilitate data sharing behavior. Funding agency and journal publisher mandates (i.e., regulative pressure) was included in the model as control variables. Results from the analysis of survey results supported TPB, with the majority of these attitudinal, normative, and control factors (all except for funding agency pressure) demonstrating an influence on data sharing behavior.

Kim and Stanton (2015) extended the TPB-based theoretical framework by considering these constructs as either individual-level factors or institutional-level factors as defined by neo-institutional theory. Attitude toward data sharing behavior is defined in the Kim and Zhang (2015) study as perceived career benefit and risk from sharing their data, and perceived effort required to share their data. The study also measured the degree to which scholarly altruism
affects data sharing behavior. Alongside these individual-level factors, Kim and Stanton examined funding agency and journal mandates (i.e., “regulative pressure”) to share data, perceived pressure from the research community to share data as part of social values and norms, and access to data repositories that facilitate data sharing as institutional-level factors. Of the individual-level factors, perceived effort and scholarly altruism were shown to be most likely to influence data sharing behavior, with perceived career benefit also having some degree of statistically significant influence. For institutional factors, normative pressure and journal publisher pressure had significant influence on data sharing behavior.

To examine the data sharing behavior of scientists working in a specific disciplinary domain, Kim and Adler (2015) analyzed a subset of survey responses collected from the Kim and Stanton (2015) study. Using the same theoretical framework and model based on TPB, the authors found some variation in the factors that affected social scientists’ data sharing behaviors. Unlike the results of the multidisciplinary respondent sample, perceived career risk was shown to be a statistically significant factor along with perceived career benefit and perceived effort. Normative pressure was the only institutional factor that was shown to have a significantly statistical influence on the data sharing behaviors of social scientists.

Zenk-Möltgen, Akdeniz, Katsanidou, Naßhoven, and Ebru Balaban (2018) also used TPB to conduct a domain-specific study of sociologists and political scientists. Similar to the Kim co-authored studies, these investigators examined attitude toward data sharing, subjective norm (i.e., social pressure to share data), and perceived behavioral control (i.e., behavioral autonomy and self-efficacy). Past behavior was also taken into account in this study’s theoretical framework as demonstration of actual behavioral control. Based on survey responses, this study arrived at results comparable to the Kim co-authored studies. Social norm, attitude toward data sharing,
and perceived behavioral control were associated with data sharing behavior, with past data sharing behavior also strongly correlated with data sharing behavior.

In a final iteration of previous analyses, Kim (2017) took into consideration the manner in which data are shared, whether in data repositories, as journal article supplements, or in response to data sharing requests via personal communication. Variations were seen in the statistical significance of the factors based on which of the three types of data sharing behaviors were reported. According to Kim, TPB “was limited in its power to explain individuals’ behavior because it does not address any contextual (such as organizational or environmental) factors influencing individuals’ behavior” (p. 881). These contextual factors to which Kim refers seem to point to actual behavioral control as a factor mediating actual behavior, just as it is modeled in TPB.

2.5 Theoretical Framework

As the aforementioned studies demonstrated, TPB offers a useful framework for investigating behavior that takes into account both the individual and contextual factors that predict behaviors and, in particular, investigating the connection between intention to perform a behavior and carrying out the behavior. As such, TPB may be helpful in explaining the dissonance between researchers’ attitudes and actions when it comes to their data sharing behaviors.

2.5.1 Theory of Planned Behavior

Azjen (1985) introduced his theory of planned behavior with a description of the banal activity of typing a letter: loading paper into the typewriter, selecting words and devising sentences, depressing the keys, etc. This set of tasks comprise a plan (whether conscious or tacit) that is requisite for accomplishing the goal of typing a letter. This illustration was to point
out the relationship between intention (goals and plans) and behavior (actions) as he formulated in the *theory of reasoned action* (TRA). According to TRA, intention to perform the behavior, which is determined by positive or negative feelings about performing the behavior (attitude toward the behavior) and social pressure to perform the behavior (subjective norm), predicts actual performance of the behavior (see Figure 3). In other words, a person is more likely to intend to perform a behavior if they have a positive attitude toward the behavior and they assume others believe they should act out the behavior. The greater the intention to perform the behavior, the greater the probability that the individual will perform the behavior.

![Figure 3. Theory of Reasoned Action model (Azjen, 1985).](image)

TRA assumes that an individual has control over their behavior, or what Azjen refers to as *volitional* control. Azjen conceded, however, that factors exist that can disrupt the intention-behavior connection by diminishing or removing volitional control. Knowledge and skill, willpower, and emotional state can have a considerable impact on volitional control. Take for instance the person who intended to quit smoking but failed to do so because of an addiction to nicotine that weakens their willpower. Along with such internal factors that affect behavior regardless of intention, external factors that affect behavior are present. Time and opportunity, and dependence on others, for example, can also preclude performance of a given behavior. A lack of sufficient time and/or the absence of opportunity will also thwart the behavior. Likewise,
if an individual’s behavior is contingent on another person’s action and that person fails to complete the action, volitional control is effectively eliminated.

The omission of non-volitional control is a shortcoming of TRA, which Ajzen remedied by modifying the framework to consider the influences on volitional control. With this modification, TRA evolved into the theory of planned behavior (TPB) that takes into account the internal and external factors that can disrupt the causal intention-behavior relationship. “Strictly speaking, intentions can only be expected to predict a person’s attempt to perform a behavior, not necessarily its actual performance” (Azjen, 2019, Author’s italics, p. 29). It is certainly possible that intention and behavior do not align due to the presence of non-volitional factors. Rather than plotting a direct line between intention and behavior, TPB seeks to understand the influences on the behavior attempts by also considering the ability of the individual to make the attempts (i.e., actual behavioral control).

TPB helps to predict behavior based on an individual’s: 1) assignment of positive or negative value toward the behavior (attitude toward the behavior); 2) belief about their peers’ approval or disapproval of the behavior (subjective norm); and 3) perception of the ease or difficulty of engaging in the behavior (perceived behavioral control). Attitude toward the behavior and subjective norm, both of which can be impacted by perceived behavioral control, determines intention, which Ajzen described as readiness to perform the behavior. Intention is manifest in observable behavior—a function of both the intention and perceived behavioral control. In short, the more positive the attitude, subjective norm, and perceived control, the stronger the intention. Given actual behavioral control, people are likely to express their intention in observable behavior. Figure 4 illustrates how the various factors predict behavior according to TPB.
2.5.2 Theory of Planned Behavior and Reproducible Research Practices

The reasons for this chasm between belief in data sharing and practice are many, but can be summarized into categories of motivating factors used in the TPB framework and referenced often in related studies: individual (attitudes and beliefs), organizational (culture and policies), and technical (resources, time and effort). Previous studies on researchers’ data sharing behaviors reached the same general conclusion that researchers are more likely to engage in data sharing practices if given sufficient incentives, mandated by policies, and provided tools and other types of data sharing support. Figure 5. illustrates the application of TPB to data sharing behavior according to those studies.

Figure 4. Theory of Planned Behavior model (Azjen, 2019).
Figure 5. TPB applied to data sharing behavior.

It should be noted here that these studies relied on surveys that collected data on self-reported attitudes and behaviors. Such surveys are susceptible to social desirability bias in which respondents feel compelled to provide answers that they believe to be more socially acceptable (Groves et al., 2009). Results of these data sharing studies showed that scientists tend to agree that sharing the artifacts of the research process should be the norm. Because key stakeholders cast data sharing as a prosocial behavior, there lies the possibility that respondents overreported this socially approved behavior, especially when considering the incongruity between: a) positive attitudes toward data sharing and the existence of data sharing policies, and b) participation in data sharing activities.

2.6 Literature Review Conclusion

In terms of the TPB model, it seems that, while researchers may have every intention to share their data and a mandate to do so, their intentions do not necessarily manifest into action. Action depends on actual behavioral control, which Ajzen (1985) defined as having the ability, resources, and any other requirements for performing the behavior. For the referenced studies on data sharing behavior, perceived behavioral control was assumed (appropriately, according to
Ajzen) to be an accurate reflection of actual behavioral control. To be sure, behavior can be performed only if the individual is capable of performing the behavior. Given that we have seen a contradiction between intention and action—even in the presence of positive attitude toward the action and subjective—a closer look at how actual behavioral control affects the intention-behavior connection is warranted.
CHAPTER 3: RESEARCH METHODS

To establish an investigative approach, the dissertation study used the theory of planned behavior (TPB) as a framework for examining the factors that may explain why authors’ attempts to share data packages that support computational reproducibility of their published results often fail. Rather than predict researchers’ data sharing behavior based on their attitudes and beliefs about the behavior as previous studies have already done, this study investigated the actual processes researchers used to prepare and share their research compendia as a means to identify the specific mechanisms that inhibited and/or facilitated their data management activities—-or in the TPB model, actual behavioral control.

To do this, I took advantage of the AJPS implementation of a data policy that requires authors to submit a research compendium that includes the data, code, and other documentation to allow a third party to reproduce the results reported in the associated manuscript prior to publication. Authors who have had an article published in AJPS since the policy took effect in 2015 presumably have performed the behaviors that yielded success in producing and sharing a reproducible research compendium. Because of this lived experience, these authors can provide the insights necessary to more fully understand researchers’ reproducible research practices (Magnusson & Marecek, 2015).

3.1 Study Design

The study adopted a mixed-methods grounded theory methodology (MM-GTM) that uses content analysis of report documents produced by verifiers to record verification results in
tandem with qualitative interviews with researchers who have experienced the AJPS verification policy.

The verification reports, which served as documentary evidence of authors’ attempts to produce and share high-quality research compendia. Verifiers use these reports to record the results of each verification attempt including specific information about research compendium deficiencies, were analyzed to identify common causes of verification failures. To further investigate researchers’ processes in preparing and submitting their compendia, I also collected and analyzed qualitative data from in-depth interviews of researchers selected from the verification report sample. Using constructivist grounded theory techniques, these interviews provided the rich data needed to identify the specific issues that impacted authors’ attempts to produce and submit a reproducible research compendium. Participants’ recollections of their experiences doing so offer insight into the first-hand challenges (or enabling resources) they encountered while attempting to perform the behaviors required to produce reproducible research compendia, i.e., actual behavioral control.

In keeping with the MM-GTM approach, I analyzed the AJPS verification reports in tandem with qualitative interviews are part of a constructivist grounded theory approach to generate an "evidence-based explanatory framework" (Creamer, 2022) that offers insights into researchers' reproducible research practices. That is, the results of the content analysis guided the structure of qualitative interviews and interpretation of interview data, while results of the qualitative analysis of interview data informed some interpretation of content analysis results.

3.1.1 Constructivist Grounded Theory

Constructivist grounded theory is characterized by concurrent, recursive data collection and analysis as part of an inductive approach for developing conceptual categories used to
explain actions and processes (Charmaz, 2014). Data collection places emphasis on “rich” data from various sources (e.g., interviews, documents, memos) based on the needs of the research question. Concurrent with data collection, analytic categories expose themselves from within the data through a coding process that synthesizes and brings meaning to the data. As categories appear, the focus of data collection may shift to unexpected areas to expand understanding of emergent category concepts. The flexible nature of constructivist grounded theory supports discovery of nuanced concepts through an iterative process of sensemaking, which makes way for the construction of theory that illustrates or explains a particular phenomenon.

In its purest application, grounded theory avoids prior assumptions about the topic of inquiry and is intended to construct new theory rather than testing or applying existing theory. Introduced in 1967 by Glaser and Strauss in their seminal text, *The Discovery of Grounded Theory: Strategies for Qualitative Research*, grounded theory was offered in response to criticism directed at qualitative inquiry. During that time, notions of objectivity, generalizability, and replicability prevailed in explications of the scientific method. This positivist epistemology gave primacy to prescriptive research methodologies for testing hypotheses by analyzing quantitatively measurable variables. Without standardized technical ‘rules’ for data collection and analysis, qualitative research was criticized as being biased and anecdotal and having little utility beyond interpretive description (Charmaz, 2014; Glaser and Strauss, 1967; Kvale, 1996). Glaser and Strauss (1967) challenged these assumptions by introducing the grounded theory approach to conducting qualitative research. This approach was meant to elevate qualitative research to the level of rigor associated with quantitative research by codifying a method for generating theory qualitatively. Grounded theory is characterized by the following techniques:
• **Constant comparison.** With constant comparison, data collection, coding, and analysis take place concurrently in a process of discovering conceptual categories and their properties as well as the relationships among those conceptual categories.

• **Theoretical sampling.** Theoretical sampling describes a data collection strategy in which conceptual categories and their relationships among one another are discovered during constant comparison, which dictates the nature and direction of subsequent data collection.

• **Saturation.** Theoretical saturation marks the point in the data collection process guided by theoretical sampling at which additional data do not offer new information about the properties of established conceptual categories.

By using these techniques, the grounded theory approach offers systematized methods wherein “dispassionate empiricism” (Charmaz, 2014, p. 9) gives way to the discovery of explanations of phenomena that are ‘grounded’ in the data.

*Constructivist* grounded theory contests the positivist nature of classical grounded theory introduced by Glaser and Strauss. Countering classical grounded theorists’ insistence on methodological objectivity and investigatory neutrality, Charmaz (2000) argued that the relative and subjective nature of knowledge construction requires a more flexible application of grounded theory methods that takes into account the role of the researcher and the study participant (and the conditions in which both are present) in *constructing* rather than *discovering* theory (Charmaz, 2014; Kenney & Fourie, 2015). Wrote Charmaz (2014):

> Hence, we constructivists attempt to become aware of our presuppositions and to grapple with how they affect the research. We aim to avoid inadvertently importing taken-for-granted values and beliefs into our work. Thus, constructivism fosters researchers’
reflexivity about their own interpretations and the implications of them as well as those of their research participants (p. 240, author’s italics).

To the constructivist grounded theorist, the ‘groundedness’ of constructed theory is not reliant on the researcher’s neutrality (which they would argue is an impossibility, and is certainly not the case in the dissertation study), but rather on the researcher’s proactive adherence to the grounded theory methods that demand analysis of observed phenomenon represented in the data while conscious of extrinsic inputs including the researcher’s a priori knowledge or assumptions (Ramalho, Adams, Huggard, & Hoare, 2015).

Again, contrary to the classical grounded theory approach defined by Glaser and Strauss that advised researchers to disregard extant theory and literature so as to allow emergence of conceptual categories from the data rather than from preconceived notions from the researcher (Glaser & Strauss, 1967; Kenny & Foury, 2015; Ramalho, Adams, Huggard, & Hoare, 2015), the dissertation study began after having established a knowledge base on the research topic from review of the literature and previous engagement with journal-based data policy implementation and enforcement, and focusing the inquiry on behavior as it is conceptualized in the existing TPB framework as a function of actual behavioral control. Constructivist grounded theory is receptive to reviews of the literature and the specification of an existing theoretical model as forming the basis of sensitizing concepts, or “points of departure” (Charmaz, 2014, p. 31), for formulating interview questions and tentative categories for initial coding of data. Data collection and analysis draw from the data to discover conceptual categories to which a path can be drawn from these points of departure while remaining open to unanticipated conceptual categories that present themselves beyond any discernible path to sensitizing concepts.
The constant comparison strategy foundational to grounded theory remains intact in constructivist grounded theory as perpetual review of conceptual categories across raw and coded data and memos take place during simultaneous data collection and analysis, with welcome disruptions of these categories as theoretical sampling explores unforeseen insights tendered by study participants until the point of saturation.

3.1.2 Constructivist Grounded Theory and TPB

It is worth noting here that TPB, as Ajzen intended, relies on quantitative survey research methods for data collection and analysis. To test the predictive power of beliefs and control on behavior, Ajzen prescribed a standard instrument comprising Likert-scale question items to elicit responses that capture the participant’s attitude toward the behavior (very good, good, bad, very bad) and their probability of performing the behavior (very likely, likely, unlikely, very unlikely). Responses to these questions are used to compute measures that represent each of the constructs of behavior in the TPB theoretical model. Regression coefficients resulting from a linear regression analysis determine the degree to which each of the factors predict the specified behavior (Ajzen, 1985; Fishbein & Ajzen, 2010).

Ajzen’s quantitative approach makes consequential assumptions about the resonance of particular aspects of the behavior to the individuals being studied, which Ajzen refers to as “salient beliefs” (Ajzen 1985). To identify salient beliefs, Ajzen advised a pilot study of a subsample of the population of interest that involves presenting open-ended questions to study participants to draw out information and terms used to denote specific behavioral outcomes (e.g., “What do you see as advantages/disadvantages of the behavior?”), normative referents (e.g., “Who are the individuals or groups that think you should/should not perform the behavior?”), and control factors (e.g., “What would help/hinder your ability to perform the behavior?”) that
are relevant to the population. A content analysis of pilot study data produces a corpus of modal salient beliefs, which are to be used to formulate question items in the main quantitative survey instrument.

It is unclear if the authors of previous studies of data sharing behavior that reference TPB as their theoretical framework executed this preliminary step of defining salient beliefs based on information provided by the population of interest prior to developing and administering the questionnaire to study participants representing the population of interest. The data sharing literature is extensive and, certainly, these authors were informed by previous studies in their selection of theoretical frameworks and definitions of sensitizing concepts. In either case, the literature presents evidence that researchers are not sharing their data despite their self-reported intentions to do, which makes salient beliefs about reproducible research practices of particular interest to the current study.

A mixed-methods research study using constructivist grounded theory provided methods that are not bound by preconceived notions of behavioral factors, and instead remained open to analytical categories delineated by the population of interest themselves beyond those salient beliefs defined in previous studies. Taken as a whole in keeping with MM-GTM, these methods helped to ensure that the construction of theory explaining a behavior is indeed reflective of the lived experiences and salient beliefs of the population of interest based on observations of their behavior (content analysis of verification reports) and expressions of salient beliefs by the individuals who performed the behavior (qualitative interviews).

3.2 Document Content Analysis

As part of the AJPS verification workflow, third-party verifiers submit verification results in a standardized form to the editor, who then send the report to the corresponding author. This
prompts the author to address all compendium deficiencies as described in the report, and to resubmit files that have been revised accordingly for a subsequent verification attempt. The verification workflow ends when reviewers find no errors in the compendium submission. At that point, the report indicates verification success, the compendium is published in the repository to allow for public access, and the manuscript is sent for publication.

Standardized verification reports are generated from a custom local database that stores verification results information that curators and verifiers enter into an electronic form. The form provides a number of fields for recording the results of the verification in two distinct sections, one for data curation and the other for code review (see Appendix A. Verification Report Sample).

The data curation section of the form includes fields for recording the list of files included in the compendium submission, results of data curation, and notes that outline issues found by the data curator that require correction. The code review section notes the technical requirements for the analysis and records the list of tables, figures, and inline results checked for reproducibility, verification results, and notes that describe code execution failures and any discrepancies between code outputs and results reported in the manuscript. Additional fields in the report input form are used by curators and verifiers for internal communications that are not included in the form sent to the editor and author for review. This internal documentation provides additional context the verification team may need to understand and address unusual situations or to store notes about the verification workflow that may be useful during subsequent verification attempts. This database contains a wealth of information for every review attempt for a given manuscript, which serves to document observations of researchers’ attempts to produce high-quality research compendia. This evidence of the performance of actual behavior,
rather than self-reported intention to perform the behavior, makes this study distinct from previous investigations of data sharing behavior.

Taking advantage of this unique source of data for the current research study, however, required the transformation of largely unstructured information into a quantifiable dataset. While the verification report form provides input fields designated for specific types of information about the verification process and results, they are designed as open-text paragraph fields that do not enforce any standard content values (i.e., controlled vocabularies) or representation rules (i.e., character formatting) to create consistencies in entries. Consequently, issues discovered during the verification process are recorded in various ways depending on how the individual curator or verifier chose to express the issue in their own words. Therefore, an analysis of the verification report documents required the use of content analysis techniques to allow for description and measurement of researchers’ reproducible research practices as recorded in the reports.

3.2.2 Document Sampling

At the time of the study, the verification process had been performed on 608 manuscripts over a period of 8 years beginning in 2015 when the AJPS data policy that required third-party confirmation of the reproducibility of reported results was first issued. Since its initial implementation, policy requirements and author guidance on policy compliance have been updated to include additional details to facilitate understanding of policy requirements. In addition, the editorship of AJPS has changed hands, with each editorial team approaching policy implementation differently and with varying levels of participation in the verification workflow.

To ensure some consistency in the study of how policy standards were applied during the verification process, the most recent (at the time of the study) 100 manuscripts that underwent
third-party review and subsequently were published in AJPS were included in the content analysis. These 100 manuscripts represented those that were subject to the AJPS replication policy in its most current form at the time of manuscript selection, with variation in the number of resubmissions that were required to successfully reproduce the results reported in the manuscript (see Table 1).

<table>
<thead>
<tr>
<th>Resubmissions</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2</td>
<td>2.00%</td>
</tr>
<tr>
<td>1.00</td>
<td>24</td>
<td>24.00%</td>
</tr>
<tr>
<td>2.00</td>
<td>39</td>
<td>39.00%</td>
</tr>
<tr>
<td>3.00</td>
<td>26</td>
<td>26.00%</td>
</tr>
<tr>
<td>4.00</td>
<td>9</td>
<td>9.00%</td>
</tr>
</tbody>
</table>

**Table 1.** Manuscripts by resubmissions.

Verification reports for each of the 100 manuscript submissions in the sample were analyzed to determine the most frequently occurring causes of verification failure and to identify the types of errors that tend to result in higher numbers of resubmissions. This began with the application of content analysis techniques to code open-ended descriptions of verification errors, which was followed by the application of a statistical model to determine the occurrence of verification errors and their impact on resubmissions.

### 3.2.3 Document Coding

Analysis of the verification report documents began with the application of an initial set of high-level codes that created distinctions among various types of errors. The initial coding process was followed by rounds of more focused coding that introduced more granularity in error
codes. I then synthesized these codes and refined them into 15 distinct error codes assigned to three categories: 1) access errors, 2) documentation errors, and 3) computational errors.

Access errors. Access to research compendium files is the minimum requirement for scientific reproducibility. To be able to verify the computational reproducibility of reported research results, the evidence base that underlies those results (i.e., data, code, and documentation files) must be made publicly accessible. Access errors were applied to verification report text segments that indicated missing files, whether due to legal restrictions or unintended omission. In addition, access error codes were applied to descriptions of issues that impede or preclude access to materials necessary to reproduce reported findings. This included problems with file formats, file quality, and technical system requirements that affected the current or future potential preservability and/or usability of research compendium files. Table 2 defines the error codes associated with physical, technical, and legal access to compendium files.

<table>
<thead>
<tr>
<th>Error code</th>
<th>Description</th>
<th>Example data</th>
</tr>
</thead>
<tbody>
<tr>
<td>accRest</td>
<td>type of access error indicating that file access is restricted due to ethical/legal issues, precluding the ability to reproduce reported findings</td>
<td>The manuscript has been granted an exemption by the editors as the [data source] data are confidential. The readme indicates that the full proprietary [data source] data is not available due to copyright issues.</td>
</tr>
<tr>
<td>accMiss</td>
<td>type of access error indicating that a file/files necessary to independently understand and retrace the analytical workflow to reproduce reported findings is/are missing from the compendium</td>
<td>Replication packages should contain a readme codebook. These files were missing in your package. Data file missing cannot run this code.</td>
</tr>
<tr>
<td>accForm</td>
<td>type of access error indicating that a compendium file was unable to be rendered properly or is stored in a file format not suitable for long-term</td>
<td>Please resubmit the codebook as a PDF for long-term preservation. There is a file [filename] in the</td>
</tr>
</tbody>
</table>
Table 2. Access error codes.

**Documentation errors.** Computational reproducibility relies not only on access to compendium files, but also the ability of researchers beyond the investigator of the original research study to understand and retrace the analytical workflow that produced the reported results. This requires that, beyond the presence of requisite files, comprehensive information describing the function of compendium files and their contents, and technical specifications and step-by-step instructions for re-executing the analysis also must be made available alongside data and code files. For compendia that omit files due to legitimate access restrictions, information on where and how to gain access to those restricted files must also be included. References to files,
dataset variables, software packages, and any other element invoked in the analytical workflow must be precisely and consistently articulated and labeled to ensure both humans and machines can comprehend and execute the analysis properly. Any text segments in the verification report indicating that documentation is missing, insufficient, inaccurate, or otherwise departing from the requirements described above are assigned a code in the documentation error category as defined in table 3.

<table>
<thead>
<tr>
<th>Error code</th>
<th>Description</th>
<th>Example data</th>
</tr>
</thead>
<tbody>
<tr>
<td>docName</td>
<td>type of documentation error indicating inconsistencies or inaccuracies in the names of files, dataset variables, or other objects in the research compendium</td>
<td>In the .do file for Table 8, there appears to be a naming mismatch for one of the .dta files…Please ensure these match. For the variables constructed from the scripts, I noticed a few discrepancies in the values/labels in the actual data versus the codebook description. Verifying the manuscript took longer than normal because several results are misidentified in code comments as reporting the wrong manuscript results.</td>
</tr>
<tr>
<td>docSys</td>
<td>type of documentation error indicating the absence, insufficiency, or inaccuracy of information about computational environment requirements</td>
<td>In README, please provide all operating system, statistical software and package information. A Mac with R version 4.0.3 was initially used for this and several coding errors occurred (see below). Running the script using the R version that is listed in the readme (3.4.2) was also attempted but similar results occurred. If version 3.5.2 is indeed necessary to run these files correctly than more information on the R package versions will be needed so</td>
</tr>
<tr>
<td>Type of Documentation Error</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td><strong>docVar</strong> type of documentation error indicating the absence, insufficiency, or inaccuracy of data variable definitions and other inconsistencies between dataset variables and variables listed in the codebook/data dictionary.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>the errors do not occur.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Please make a note in the 'readme.txt' file that the analysis preparation file (&quot;[filename]&quot;) must be used on a Mac in order for it to run correctly.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>In Codebook:</strong> all values must be labeled for all variables, please review codebook to ensure that all values are present for any binary, dichotomous, dummy, and/or categorical values.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Codebook must be a 1:1 match with variables in dataset. There are quite a few variables in the data that are not represented in the codebook. If these are not necessary for the analysis, they can be removed from the data. If they are required for the analysis, they must be included in the codebook with variable names, definitions, values, and value labels (including missing values).</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>docAcc</strong> type of documentation error indicating the absence, insufficiency, or inaccuracy of information about data sources (i.e., data citation) and how to access the specific datasets used in secondary analyses</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full data citations for all original source data must be provided in either the README or Codebook.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proprietary data access requirements must be described in the README -- it is not enough to link to their main site. Please provide secondary users instructions for requesting access to the exact data you used in your analyses.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>docRead</strong> type of documentation error indicating missing, inaccuracy, or insufficient descriptions of submitted research compendium files and their functions in the computational workflow</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All files from the Dataverse must be listed and described in the README. There are quite a few files missing from the README in the ado section, please update to meet this requirement.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
If we were to request one small change, it would be to clarify in the readme file which scripts produce which manuscript results, which would make partial/focused replications of the results easier (because replicators wouldn’t need to open every script and inspect the code to see what it produces).

Table 3. Documentation error codes.

*Computational errors.* Even with access to a complete and well-documented compendium, claims of computational reproducibility can be made only if the analysis was re-executed successfully and without errors, and the outputs of the re-executed analyses are identical to those presented in the published reports. Whenever this is not the case for submitted compendia, verifiers described the various reasons that caused reproduction failure in the verification report. These reasons may have included omission of requisite code commands as well as coding issues ranging from basic syntax errors to missing declarations of software dependencies that triggered errors or use of file paths pointing to directory structures exclusive to the original investigator’s unique computational environment. In cases in which code ran successfully, computational error codes were still applied where the report indicated that the computational workflow was not coherent due to poor coding practices or the lack of non-executable comments describing the workflow, or if outputs did not match reported results. Definitions and examples of these computational errors are presented in table 4.
<table>
<thead>
<tr>
<th>Error code</th>
<th>Description</th>
<th>Example data</th>
</tr>
</thead>
<tbody>
<tr>
<td>compExec</td>
<td>type of computational error indicating failure of the code to compile/execute successfully</td>
<td>For a reason I can’t quite figure out, I wasn’t able to execute the entire .do file in a single pass. I kept receiving “invalid ‘x’ r(198);” in the Appendix Table 4 codeblock.</td>
</tr>
<tr>
<td>compOut</td>
<td>type of computational error indicating the mismatch between outputs generated from the submitted code and data and outputs reported in the manuscript</td>
<td>The code compiled successfully but the figure did not fully match the version in the main text. I am unsure what’s going on, perhaps a difference in seed randomization?</td>
</tr>
<tr>
<td>compDir</td>
<td>type of computational error indicating failure of the code to account for the directory structures of an independent computational environment</td>
<td>Within all R scripts, please remove personal file paths from setwd() calls and include the readme comment in the script reminding users to set the working directory. Please ensure calls to data and R scripts within the analysis reflect the structure of the Dataverse files. Ie either include dta/code/results sub-folders, or remove reference to them within the analysis.</td>
</tr>
<tr>
<td>compDoc</td>
<td>type of computational error indicating the absence, insufficiency, or inaccuracy of non-executable code comments that describe the function of code commands</td>
<td>Given the structure of your code, please provide a comment in the working directory that users should run [filename] and [filename] from within [filename],</td>
</tr>
</tbody>
</table>

Model 1: ...standard errors should be .02, which likely will change statistical significance (reported as .04).
so they will be prompted to install and load necessary packages for each.

The code files were quite difficult to follow and seemed to include quite a lot of unnecessary comments and code. For the resubmission please take effort to clarify the code.

For R code, please provide “install.packages” commands to prompt users to install necessary packages.

I was able to reproduce this, but I had to manually install an old version of interfex. I recommend either requiring that users install the version you have used or providing a comment that it may require your exact version in order to replicate.

Table 4. Computational error codes.

The error codes applied to text segments of the verification report documents allowed for the quantitative analysis of the outcomes of the authors’ attempts to produce and share high-quality research compendia.

3.2.4 Statistical Analysis

The analysis used ordinary least squares (OLS) linear regression to determine the types of errors that influence the number of times an author was required to resubmit their research compendium before it was successfully reproduced, with the number of resubmissions as the dependent variable, and the individual error types as independent, or explanatory, variables. The OLS regression was performed using the equation

\[ Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + e_i \]

whereby \( Y \) is the dependent variable, each \( X \) is an independent variable, and \( e \) represents
unobserved variables that affect $Y$ (i.e., stochastic error term (Gujarati & Porter, 2009)). $\beta_0$ is the constant coefficient, or the value of $Y$ when the independent variable is equal 0. The value of $\beta_1, \beta_2, \ldots, \beta_k$ is the regression coefficient for $X_1, X_2, \ldots, X_k$, respectively. A positive regression coefficient indicates an increase in the dependent variable when the independent variable increases; a negative coefficient value indicates a decrease in the dependent variable when the independent variable increases.

In terms of the current analysis, the OLS regression equation was constructed where the dependent variable was defined as the number of resubmissions required for successful verification, and the independent variables as dummy variables representing each of the three types of errors found to be present in the compendium: access errors, documentation errors, and computation errors. Thus, the OLS regression equation used in the analysis was as follows:

$$\text{Resubmissions}_i = \beta_0 + \beta_1 \text{AccessErrors}_i + \beta_2 \text{DocumentationErrors}_i + \beta_3 \text{ComputationErrors}_i + e_i$$

The OLS regression was calculated in R using the `lm()` function, with results showing the degree to which independent variables (i.e., error categories) have a statistically significant effect on the dependent variable (i.e., number of resubmissions).

Based on findings from the content analysis, a subset of cases was selected for qualitative interviews where independent variables show a higher statistically significant effect on the number of resubmissions. Selection of the interview sample also considered the diversity of cases to include those with few (two or fewer) resubmissions and those with many (three or more) resubmissions.

3.3 Qualitative Interviews

The qualitative interviews were conducted using data collection and analysis techniques from constructivist grounded theory to learn more from authors about the granular processes they
used to successfully fulfill the verification policy. This includes collecting data about the
challenges they encountered, and any resources used during the course of preparing their
research compendium for submission and review. These techniques focused on concurrent,
recursive qualitative data collection and analysis as part of an inductive approach for developing
conceptual categories used to explain actions and processes (Charmaz, 2013).

3.3.1 Participant Sampling

From the 100 verification reports analyzed, corresponding authors were selected to
participate in an interview to discuss their experience with the AJPS verification process. To
ensure sample diversity, interview participants included authors with varying degrees of success
in producing a reproducible research compendium based on the number of times they were
required to resubmit materials as a result of compendium deficiencies. Sampling also considered
results of the content analysis, ensuring that the final sample included authors whose research
compendia were found to have errors that the content analysis found to have a statistically
significant impact on the number of resubmissions.

The number of interviews that were conducted was dependent on the degree to which
additional interviews yielded no new insights into conceptual categories and offered no
additional information that could be used to characterize the categories. As part of this
theoretical sampling strategy, interviewing ceased after an assessment of the persistence of
conceptual categories established during initial and focused coding, which indicated saturation of
categories for which new data failed to reveal new theoretical directions or patterns in the data,
leaving only repetition of established codes. Because saturation is a necessary component of
grounded theory for demonstrating robustness of theoretical categories, the study was prepared
to interview approximately 20 participants to ensure that the data collected are sufficiently rich to
achieve theoretical category saturation (Guest et al., 2006). This number was an estimate based
on studies of sample size sufficiency for thematic saturation—which provide vastly different recommendations based on very different methods for determining sample size sufficiency (Crouch & McKenzie, 2006; Fugard & Potts, 2015; Guest et al., 2006; Magnusson & Marecek, 2015). Adhering to this standard, theoretical saturation was achieved with 14 interviews. Table 7 shows the characteristics of the selected interview sample.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assistant</td>
<td>8</td>
<td>57.1</td>
</tr>
<tr>
<td>Associate</td>
<td>3</td>
<td>21.4</td>
</tr>
<tr>
<td>Lecturer</td>
<td>1</td>
<td>7.1</td>
</tr>
<tr>
<td>PhD Candidate</td>
<td>2</td>
<td>14.3</td>
</tr>
<tr>
<td>Total</td>
<td>14</td>
<td>100.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Resubmissions</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>14.3</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>42.9</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>35.7</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>28.6</td>
</tr>
<tr>
<td>Total</td>
<td>14</td>
<td>100.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Errors</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access errors total</td>
<td>12</td>
<td>85.7</td>
</tr>
<tr>
<td>Documentation errors total</td>
<td>13</td>
<td>92.9</td>
</tr>
<tr>
<td>Computation errors total</td>
<td>13</td>
<td>92.9</td>
</tr>
</tbody>
</table>

Table 5. Interview sample characteristics.
3.3.2 Qualitative Data Collection

To elicit rich data from interviews, open-ended questions were designed to encourage participant-centered conversation about their process of assembling and submitting their research compendium. This intensive interviewing technique encouraged participants to share specific details about their perspectives and behaviors, with as few constraints as possible on participants’ natural inclination to share information they appraise as being relevant to them personally. An interview guide established the context and content of the discussion and ensured that the information provided by participants was sufficient enough to allow for the emergence of analytic categories. However, questions were not limited to those in the interview guide; I introduced additional probing questions at times during the interviews to draw out additional or clarifying details or to pursue unforeseen, but relevant, concepts introduced by participants.

Questions were formulated using sensitizing concepts based on the original TPB framework, which has been used in previous studies of data sharing behavior and tested in many disciplinary domains. Additional sensitizing concepts were drawn from the content analysis of verification reports, which identified issues in authors’ attempts to produce and share their research compendia. The interview guide (see Appendix B. Interview Guide) was subject to a pretest in which individuals with some knowledge of sensitizing concepts participated in trial runs of the interview. Pretest interviews informed necessary changes to the content and delivery of interview questions that resolved issues of ambiguity or confusion, ensured that questions elicit the rich data required for analysis and theory formulation, and assessed the feasibility of the interview duration.

Invitations were delivered to potential respondents via email requesting their participation in the study (Appendix C. Study Invitation). Along with information about the purpose of the
study, the invitation also described the nature of their participation in terms of the interview content, estimated duration, and platform. All other IRB-required study information and informed consent language were included in the invitation text (Appendix D. Informed Consent).

Interviews were conducted using Zoom web-based conferencing software that supports end-to-end audio and video encryption. Prior to their interview sessions, I asked researchers to review a formal consent form and confirm their consent to participate in the study in accordance with the language in the consent form. The consent form specified the purpose of the study, the nature of their participation in the study, how the information they provide will be used in the current and future studies, the potential benefits and harms that may come about from their participation, and other information required by the UNC Institutional Review Board to satisfy informed consent requirements for the conduct of research involving human subjects. With confirmation of the participant’s consent to do so as declared in the consent form, I recorded and transcribed interviews for qualitative analysis. Once necessary changes were made, interview recordings were deleted and no longer available for access and review.

3.3.3 Qualitative Data Analysis

Data analysis occurred concurrently with data collection so that subsequent interviews were able to take into consideration revelations from previous interview data, which prompted adjustments to the interview guide for subsequent interviews. Introducing new questions or question probes in later interviews enabled further exploration and understanding of new potential conceptual categories. Despite participants who interviewed prior to interview guide updates not having been presented with the new questions, the goal of theoretical saturation remained, requiring arrival at the point of data collection in which no new information about a conceptual category emerged. Data analysis included a rigorous coding process that involved both initial coding applied during line-by-line review of interview transcripts followed by
focused coding that refined and condensed initial codes into abstract conceptual categories for analysis and interpretation. During simultaneous data collection and coding, constant comparison among codes and categories across interview data captured in analytic memos. This advanced the analysis toward theoretical saturation and the development of new theory in the form of a narrative construct that offers interpretation of the influence of actual behavioral control on the intention-behavior relationship.

ATLAS.ti qualitative analysis software was used for coding, memo-writing, and analysis. ATLAS.ti offers functionality for constructing codes and applying them to interview transcript and memo text segments and creating code groups that support focused coding and sensemaking. Additional tools for searching and comparing across data, codes, and memos facilitated the recursive, iterative nature of grounded theory approaches for data collection and analysis.

During a line-by-line review of the interview data as part of an initial, or open, coding process, code labels were applied to segments of interview transcript text that summarized general observations and distinguish data fragments—with little respect to preconceived or anticipated ideas about the data beyond sensitizing concepts. In addition, the line-by-line review was used to discover specific terms or phrases introduced by participants during their descriptions and recollections. These terms and phrases were used as in vivo codes to signal areas that required additional exploration and to understand their implicit meaning within the context in which respondents expressed them. Focused coding refined and condensed initial codes to identify possible theoretical insights that established the analytic frame of the study.

Following a critical review and comparison of initial codes and the data fragments to which initial codes are applied, substantive focused codes derived from initial codes were used to further synthesize the data for explicit analysis and theory construction. Codes that reappeared across interview instances and/or that offered particular clarity about an area of focus were given
the status of provisional theoretical categories that warranted further probing in subsequent interviews.

Throughout data collection and coding, memos were used to capture thought processes that underlie chosen codes and conceptual categories, articulate substantive questions that arose during data collection and analysis, and recorded any other ideas or conjecture about the data and codes. Memo-writing is an exercise of “critical reflexivity” (Charmaz, 2014) that reinforces emphasis on the data in the analysis and theory construction process. While informal, these analytic notes also serve as a record of the research workflow and the factors that led to decisions about selecting and applying codes, and articulate the basis for comparisons drawn between interviews, codes, and categories. As more data were collected and coded, memos also facilitated sorting and integration activities that led to identification of both emergent categories and analytic gaps that require reexamination of data or collection of additional data.

Initial and focused coding were considered tentative with modification permitted as new data were collected and reviewed alongside data segments already tagged with codes. Any strong notions of analytic categories that arose during initial coding were documented in accompanying memos for later consideration or questioning when attempting to make associations with empirical observations and contextual analysis during focused coding. These considerations and questions prompted reassessment of present data and/or collection of additional data to delineate or expand meaning of conceptual categories as part of the theoretical sampling strategy. This promoted continuous development of categories to understand their properties, context, and relationships to other categories to support reasonable inferences about empirical observations captured in the interview data (Glaser & Strauss, 1967).
3.4 Study Validity

While qualitative research methodologies are not as conducive to measures of validity and reliability as are quantitative studies, Sikolia, Mason, Brios, and Weiser (2013) noted their equivalents in grounded theory research. According to them, trustworthiness of research results from grounded theory studies, characterized by credibility, transferability, dependability, and confirmability, can be established through data triangulation, participant confirmation of interview transcript content and emerging conceptual categories, comprehensive documentation of research processes, and tests of intercoder reliability, respectively. These strategies were applied as described below.

Data triangulation. The analyzed data from participant interviews alongside verification reports that provided documented evidence of participants’ adherence or non-adherence to the journal policy. While interview data captured the salient beliefs of the participants and recall of their process of preparing materials for submission, the verification reports confirmed participants’ recollections of their experiences with the data policy, while also informing the selection of a sample composed of participants with varying experiences.

Documentation. Throughout the research process, research plans and processes were documented in a methodological journal. Any complications that arose and decisions made as a result of these complications were also recorded. In addition to the methodological journal, analytic memos were written during initial and focused coding to provide specific details on the rationales behind establishment and application of analytic codes and categories. The journal and memos made the research process more transparent, allowing for the possibility that others can repeat the study.

Intercoder reliability. To ensure that conceptual categories were indeed derived from the data, the study performed an intercoder reliability assessment. During the coding phase of the
study, an individual with knowledge of sensitizing concepts was enlisted to apply initial codes to a subsample of the data. The alignment between the two instances of code application signaled the appropriate specification of codes to the data beyond the viewpoint of the principal investigator.

3.5 Role Considerations

The AJPS is under contract with the Odum Institute for Research in Social Science to perform independent verification of accepted manuscripts, which is managed by Odum Institute Archive staff. As the assistant director for the Archive, I have responsibility for overseeing verification workflow activities, which includes privileged access to information from non-public verification reports containing details of verification attempts for each manuscript that has undergone the verification process.

Anticipating its research value, this information has been and continues to be standardized and stored in an electronic database. The AJPS editors have granted permission to use this data for the purposes of this and related research. The only use of this privileged information was for the content analysis and selection of potential interview respondents. Contact information for respondents was accessed using the author information provided with the published article. All IRB requirements were followed throughout the study.
CHAPTER 4: RESEARCH RESULTS

This chapter describes study results from the content analysis of verification reports and qualitative analysis of author interviews. Content analysis results identify the issues that more significantly impact the reproducibility of research compendia. Results from the qualitative analysis offer additional evidence of the factors that influence researchers’ reproducible research practices while also introducing new insights into the actual behavioral controls that affect researchers’ success in producing research compendia that meet quality standards for reproducibility.

4.1 Content Analysis Results

The coding and content analysis of verification reports identified the most frequently occurring errors found in submitted research compendia. Table 6 presents the number of compendia that were found to have the error type listed.

<table>
<thead>
<tr>
<th>Error Type</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access errors</td>
<td>72</td>
</tr>
<tr>
<td>Documentation errors</td>
<td>98</td>
</tr>
<tr>
<td>Computation errors</td>
<td>87</td>
</tr>
</tbody>
</table>

Table 6. Manuscripts by error type.

The error frequencies show that documentation errors (n = 98) were the most often found by verifiers, followed by computation errors (n = 87) and access errors (n = 72).

Concerning the impact of these error types on the number of resubmissions required for successful verification, results of the OLS regression analysis show computation errors to have
the highest impact on resubmissions. The analysis results also show a positive association between access errors and number of resubmissions, but to a lesser degree than computational errors. Despite documentation errors being the error type that is most often found during verification, it did not have a statistically significant effect on resubmission numbers. Table 7 presents the results of the OLS regression.

<table>
<thead>
<tr>
<th>Factors/Independent variables</th>
<th>Coefficient</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access errors</td>
<td>0.25091*</td>
<td>0.10411</td>
</tr>
<tr>
<td>Documentation errors</td>
<td>0.11814</td>
<td>0.08197</td>
</tr>
<tr>
<td>Computation errors</td>
<td>0.18714**</td>
<td>0.07039</td>
</tr>
<tr>
<td>Error categories</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.2587</td>
<td></td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>0.2355</td>
<td></td>
</tr>
<tr>
<td>F-statistic</td>
<td>11.17***</td>
<td></td>
</tr>
</tbody>
</table>

*p < 0.05. **p < 0.01. ***p < 0.001.

Table 7. Results of OLS regression.

The R-square of .2587 ($F = 11.17, p < 0.001$) indicates that the independent variables account for 25.87% of the variation in the dependent variable. In other words, 25.8% of the variance in the number of resubmissions can be explained by the presence of the errors.

During interviews in which they described the difficulties they encountered when preparing and submitting their compendium files for verification, researchers offered some explanation as to why some types of errors had a greater impact on the number of resubmissions. The following section describes findings from those interviews.
4.2 Qualitative Interview Analysis Results

The qualitative analysis of interview transcripts identified several factors that corroborate findings from previous studies that found that attitudes toward data sharing (incentives), subjective norms (policies), and perceived behavioral control (support) influence data sharing behaviors, as the theory of planned behavior asserts. Beyond this confirmation, additional conceptual categories emerged from the analysis that offered some explanation as to why, despite the presence of those factors that promote data sharing behaviors, scientists struggled to engage in the behaviors that support research reproducibility. In the following sections, I describe both the known factors that contribute to data sharing as well as those that address the discrepancy between data sharing attitude and data sharing action.

4.2.1 Beliefs About Reproducible Research

This section outlines the TPB factors affecting reproducible research behaviors based on participant responses to questions that elicited their perspectives on the AJPS data policy, its goals, and the value of policy implementation. Examples of these responses from the data are included to further illustrate these themes relevant to beliefs about reproducible research practices that influence behaviors corresponding with those beliefs.

Behavioral beliefs (attitude toward the behavior). Overall, participants were largely supportive of the policy and its goals and were willing to perform the actions required to uphold policy standards for reproducible research compendia. For the most part, participants—even those who described the process of assembling and submitting their compendia for verification as labor intensive and cumbersome—expressed their appreciation for the quality standards applied during the verification process and the value of providing access to high-quality reproducible research compendia to the research community. Moreover, none of the participants
found the policy to be a deterrent, given the career incentives that come with publishing in a high-impact journal. Particularly for junior faculty, publication in the AJPS is highly desirable for tenure and review, thus increasing their favorability of the policy. Table 8 describes these behavioral beliefs and the largely positive attitudes toward reproducible research practices.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Beliefs</th>
<th>Sample data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Behavioral beliefs</strong></td>
<td><strong>Attitude toward the behavior</strong></td>
<td></td>
</tr>
</tbody>
</table>
| 1 Support for policy goals    | Fulfilling policy requirements supports research reproducibility | *I think I am generally supportive of the verification policy. I thought it was useful for my own publication in the sense that I was basically forced to work through the code again and make sure that it works and looks somewhat readable. And I think also for research in general, right, there are many issues with reproducibility, so I guess I think it’s a step in the right direction.*  
*I think it’s good. I found it onerous in a way to go through the data verification process. But at the same time I’m grateful to have gone through it because I think it does, you know, hold the research up to kind of the highest standard of making sure that everything is completely verifiable.*  
*I think it’s an important step for transparency and social science, like, I think it’s a valuable thing. I think that the ultimate replication packages that come out of it are much easier to use both as a researcher and as a teacher, like, as a professor teaching quantitative methods…*  
*It can get messy while the project is ongoing, and so I feel like packaging everything for the replication archive is a good way of sort of cleaning up the
project a bit and organizing, because, you know, definitely, like the structure that AJPS had us do is better for a finished product.

I submitted, and like, based on AJPS’s reputation as a journal, I just knew that, like, if I was fortunate enough to be able to publish that I would have to make sure things were replication ready.

I would say no [that I considered submitting my manuscript elsewhere because of the rigor of the policy] because of the prominence of AJPS in the discipline. I could imagine, maybe in the case of a, you know, lower tier journal… But in the case of AJPS, just given its prominence, you know, I was all in. All the times I’ve submitted there, I’ve sort of been ready and willing to do the replication process if it comes to that.

I’m not aware of all of the policies of other competing journals. Let’s say I see one of the best. So, in a way, it’s a burden that is underestimated, I think, at the beginning. But it’s still something that perhaps especially junior faculty are keen on taking on because of the career kind of concerns and what the payoffs are in terms of, yeah, getting a job.

Table 8. Behavioral beliefs affecting reproducible research practices.

Normative beliefs (subjective norms). As previous studies found, regulatory pressure is a factor that has a statistically significant effect on data sharing behavior. In the case of the participants of this study, all were subject to the regulatory pressure of the journal policy and performed the behavior of interest accordingly. Because publication in AJPS is contingent on successful verification, failure to meet policy requirements for reproducibility was not an option for them (except in rare cases in which the editor granted a policy exemption because verifiers
were not permitted or able to access materials due to legal or ethical reasons). Even so, several participants noted that making their research compendium accessible has become a community expectation, with some of them noting that they had already made it their practice to engage in reproducible research practices and continue to do so regardless of whether or not a journal policy requires them to do so. Table 9 presents the normative beliefs and the subject norms described by researchers during their interviews.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Beliefs</th>
<th>Sample data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Normative beliefs</strong></td>
<td><strong>Subjective norm</strong></td>
<td></td>
</tr>
<tr>
<td>1   Policy requirements</td>
<td>Failure to fulfill policy requirements is not an option.</td>
<td>Unfortunately, AJPS is kind of a career changing place. Um, it has a humongous responsibility, and [paper rejection due to verification failure] would be a major cause for panic if I had been untenured.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>As an author, at least me, I can’t really be choosy about where I send stuff, so I think the value of an AJPS still outweighs the burden.</td>
</tr>
<tr>
<td>2   Community expectations</td>
<td>Policy requirements align with community expectations.</td>
<td>Experience has taught me the importance of just being really, you know, having better systems, and, you know, informative file names like, you know, leaving better code, just all of those things. And I think seeing a rising standard within political science over the past decade in terms of expectations of transparency, I think, has just encouraged me to keep better records than I did before.</td>
</tr>
</tbody>
</table>
I think I’m not sure it’s specific to the experience of AJPS, or that there’s just an understanding when you start a project to know that you’re going to need code to produce the final tables and figures when you’re done with it. So I, as part of that kind of, like, inculcation of that lesson that, yeah, everything has to be perfect, and if it’s not, then you’re doing it wrong and you’re going to make a lot of trouble for yourself later.

Table 9. Normative beliefs affecting reproducible research practices.

*Control beliefs (perceived behavioral control).* Computation-based analyses require a baseline of technical skills to write programming scripts that execute properly and produce plausible results, which all participants eventually demonstrated by having had their compendia verified and confirmed to be reproducible (again, except for those who received policy exemptions for legal or ethical reasons). Thus, when initially confronted with policy requirements, participants expressed confidence that verification of their research compendia would result in success. For them, it was unexpected that verifiers would find deficiencies or errors in their compendia that would require correction (which is the case for the vast majority of submissions). Some researchers noted that their experiences submitting to other journals with a data policy in place informed their expectations of the AJPS verification process, which, as they discovered, had more rigorous requirements than other journals’ data policy requirements. Table 10 presents the control beliefs and corresponding perceptions of behavioral control synthesized from researchers’ interviews.
<table>
<thead>
<tr>
<th>Factor</th>
<th>Beliefs</th>
<th>Sample data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control beliefs</td>
<td>Perceived behavioral control</td>
<td>I understood the goals of the policy were to facilitate replication and transparency, um, and so I guess I assumed it would be similar to the experiences I had with other journals in the past…</td>
</tr>
<tr>
<td>Existing technical skillsets</td>
<td>Skillsets are sufficient to fulfill policy requirements.</td>
<td>I mean, honestly, my expectations [about the policy requirements] were, I know this all works on my computer, so I should be able to more or less just ship all the files over the Odum and tell them where to change the working directory, and that’s about it.</td>
</tr>
<tr>
<td>Prior experience</td>
<td>Previous experiences informed strategies for fulfilling policy requirements.</td>
<td>It’s been part of my practice since the beginning to always have reproduction archives for all of my papers…That’s just to say, like, I’m glad that [AJPS was] bringing everyone else up to the transparency goals that I set for myself. Experience has taught me the importance of just…having better systems and, you know, informative file names like, you know, leaving better code. Just all of those things. And I think seeing a rising standard in political science over the past decade in terms of expectations of transparency, I think, has just encouraged me to keep better records than I did before.</td>
</tr>
</tbody>
</table>
According to the TPB framework, the presence of these belief factors should be predictive of intention to perform the behavior associated with those beliefs. Given that the vast majority of researchers who submitted manuscripts to AJPS after the verification policy was issued (despite positive attitudes toward policy requirements, pressure from subjective norms, and perceived behavioral control), actual behavioral control becomes a critical area for investigation when failure to perform the required behavior is commonplace.

### 4.2.2 Reproducible Research Behaviors

TPB posits that an individual’s intention to perform a behavior becomes observable behavior only if the individual has actual behavioral control. During their interviews, participants were asked to recall their process of assembling and submitting their research compendium, and the challenges they confronted during their attempts to do so. This focus on their specific activities captured information on actual performance of the behaviors of interest. Conceptual categories that emerged from the qualitative analysis of the interview transcripts were centered on the TPB element of actual behavioral control. Interpretation of these conceptual categories yielded factors for actual behavioral control that describe in more granularity the direct impacts on researchers’ ability to produce reproducible research compendia. These factors are described in table 11.

<table>
<thead>
<tr>
<th>Factors</th>
<th>Behavioral controls</th>
<th>Sample data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Behavior performance</strong></td>
<td><strong>Actual behavioral control</strong></td>
<td></td>
</tr>
<tr>
<td>1 Acceptance of policy conditions</td>
<td>The ability to fulfill policy requirements is assumed.</td>
<td>-</td>
</tr>
<tr>
<td>2 Comprehending policy requirements</td>
<td>Policy requirements are not explicit enough to enable</td>
<td>So, what I would have wanted is a much clearer, more specific</td>
</tr>
</tbody>
</table>
guide because there’s, like, a lot of very specific standards that are enforced...And so, for example, we had, you know, binary variables like ‘0’, ‘1’. We just said ‘indicator for x’ right? And we got back, and they were, like, for each of these you need to say what ‘1’ is and what ‘0’ is. And, sort of, if we had known that up front, we could have done that the first time and reduced the, sort of, amount of work that you had and the amount of work we had.

It was just my lack of knowledge, um, that was the roadblock. I just didn’t realize or didn’t know, um, you know, that these small details about descriptions of variables in a codebook, and what the readme files should look like, and everything else like that. Um, I guess I just didn’t really realize that they really meant that, you know, that really needs to be there. And so if I had just, um, come to that realization more quickly, I think it would have been a pretty fast process that time around.

3  Encountering technical limitations

Technical limitations present obstacles to fulfillment of policy requirements.

But we did run into a problem, which was that a particular R package had been updated at some point between the time when we first did the analyses, and by the time we were going through the replication process, and it didn’t make any major changes to the research. You know, this changed that, like, the third decimal point in a table of results. But, of course, that was a problem...So we had to do a
lot of back and forth that time to figure out exactly which package it was that was causing problems and then come up with a solution, which was to use a particular package from the archive, or a particular version of that package from the archive. But that took a while to get to that point.

I would say that there was this kind of big picture issue that was kind of complicated to deal with, which was how to handle the replication and data posting and verification for data that did not originate from us, the researchers, and therefore there were restrictions on posting it.

Table 11. Actual behavioral control factors affecting reproducible research practices.

Aspects of actual behavioral control identified in the data suggest that the difficulties researchers experience during their process of producing and sharing reproducible research compendia are not a result of their inability to do so (i.e., volitional control). Rather, actual behavioral control is diminished when policies fail to clearly and explicitly articulate expectations and no models exist to illustrate what fulfillment of policy expectations looks like, when access to compendium components is restricted for legal or technical reasons, or when the idiosyncrasies of technology stymie third-party reproduction attempts.

Acceptance of policy conditions. The AJPS policy was issued without regard for whether or not researchers are capable of fulfilling policy requirements. Indeed, all researchers who had an article published in AJPS since the policy took effect have succeeded in producing and sharing a reproducible research compendium. Based on this fact, it is reasonable to assume that
they possessed the ability to perform, at least eventually, the activities required to meet the conditions of the policy.

*Understanding of policy requirements.* Researchers have the capacity to fulfill policy requirements, but only if the policy provides sufficient information for researchers to perform the necessary actions to meet expectations. Researchers noted that AJPS policies are exacting, with specific parameters for documentation contents and file formats and coding practices. Without these parameters being apparent to researchers, they cannot be expected to abide by them. For many researchers, reproducible research practices were not taught as part of their graduate training when they were learning computational analysis techniques. Instead, knowledge of these practices came from prior policy compliance experiences or from observing the practices of their peers. Community-based quality standards for producing high-quality research compendia have yet to be established, so researchers rely on the policy to articulate the standard. But without detailed guides or models that represent research compendia that meet policy expectations for high-quality compendia, some researchers struggled to fulfill policy requirements.

*Encountering technical limitations.* Access to the data, code, and documentation needed to understand and retrace the computational workflow of an analysis is the minimum requirement for reproducibility. When access to compendium files is restricted due to legal or ethical issues, verifiers must seek provisions for special access to the restricted materials. However, these provisions are not always available nor feasible. This is a technicality that prevents verifiers from performing the full verification workflow. Another obstacle that hinders successful verification is the reliance of computational analysis on the particularities of hardware and software. Even minor differences in hardware, operating systems, software packages can cause notable differences in analysis outputs. Researchers described various instances of these
technical limitations that had to contend with before their manuscripts were granted final acceptance and published.

These are non-volitional factors that must be addressed to support researchers’ success in performing the behaviors required to produce and share reproducible research compendia. When these issues are resolved, researchers are able to successfully produce and share reproducible research compendia. Eliminating the presence of these non-volitional controls in the first place is likely to reduce the number of resubmissions, which will lower the cost of policy implementation, thus making it more feasible to a greater number of journals.

One other finding of note from the qualitative analysis was the positive effect that policy compliance had on study participants’ research workflows. Having gone through the verification process, several researchers indicated that they have made changes to their data management practices since they published in AJPS. In response to an interview question that asked participants if and how their research practices after having gone through the verification process, one researcher explained:

Yeah, I would say so. In fact, I even say this to students, too. Like, you need to be writing your code with the idea that this is going to end up in a replication archive someday, and I try to follow my own advice on that, too, and really, really be clear, at least in the documentation...definitely in how I write the code and comment and just being aware of package versions I’m using, all of that has increased quite a lot since I’ve been submitting replication data to journals.

Beyond ensuring that research compendia meet quality standards that allow for validation of published scientific results, researchers indicated that the AJPS policy has had a positive impact on their normative research practices to include data management activities that support
reproducibility. This finding, however, is based on self-reports and requires further study for substantiation. Still, knowledge of the non-volitional controls that inhibit reproducible research practices is useful for improving the efficacy of journal-based policies. At least in the specific instance of policy compliance, researchers are compelled to understand and execute reproducible research practices, giving them the knowledge and experience that is needed for adoption.
CHAPTER 5: DISCUSSION AND CONCLUSION

Study results pointed to the impact of the implementation of a journal-based policy that obligates authors to engage in the activities required to make their research reproducible. A strict policy that makes publication contingent on compliance, clearly articulates its requirements, and accommodates technical limitations mitigates the impact of non-volitional controls to allow researchers to perform the desired behaviors promoted by the policy.

5.1 From Policy to Practice: Considering Actual Behavioral Control

Based on the results of the content analysis and the qualitative interviews, I propose the conceptualization of actual behavioral control as comprising three “sub-controls”: ultimatum, articulation, and accommodation. These sub-controls, which can be aligned with TPB belief factors, are the mechanisms that support policy compliance, which in turn promotes adoption of reproducible research practices. Figure 6 illustrates these sub-controls as components of actual behavioral control within the TPB model. Policies that consider these actual behavioral sub-controls in their implementation increase the impact of actual behavioral control to close the gap between beliefs about reproducible research practices behaviors and actual performance of reproducible research behaviors.
Figure 6. Components of actual behavioral control in the context of journal-based data policy compliance.

5.1.1 Ultimatum

In the case of the researchers subject to the AJPS verification policy, performing the behaviors required to fulfill policy requirements is not optional. The policy presumes that the researcher has the actual behavioral control to do so. Thus, the intention construct in the TPB model is replaced with policy, as submitting a manuscript to AJPS demonstrates acceptance of the terms of the policy, which denotes intention to carry out the behaviors required to fulfill policy terms.

Indeed, every article published in AJPS since the verification policy took effect has successfully undergone the verification process. This fact serves as evidence of researchers’ ability to successfully fulfill policy requirements even when faced with obstacles (volitional or non-volitional) that may have made it difficult to do so. When a journal-based data policy presents an ultimatum, researchers will find ways around obstacles to perform the
required behaviors in order to earn the incentives that publication in a top-tier journal offers.

5.1.2 Articulation

Researchers indicated that their knowledge of reproducibility standards for research compendia was limited to their experiences submitting to journals with less rigorous policies or from some awareness of community expectations for reproducible research. These prior experiences and awareness of standards did not prepare researchers for the rigor of the AJPS verification policy. Without clear guidance on how to meet the standards of the AJPS policy or access to exemplar compendia that illustrated the standards, researchers submitted substandard compendia that required modification and resubmission to pass verification.

Additionally, if policies are to serve as a primary mechanism for enforcing and promoting the adoption of reproducible research practice, then policy requirements must be intentionally aligned with the goals of the policy and practices to further support understanding and acceptance. To reduce the number of resubmissions, and thus the amount of time and costs of implementation, a journal-based data policy must articulate its requirements clearly, with guidance materials, training, and exemplars that represent a range of analytical workflows.

5.1.3 Accommodation

Both the results of the content analysis and the interviews showed that technical obstacles and technicalities stood out as a hindrance to policy compliance. In many cases, these obstacles are no fault of the researcher. Instead, idiosyncratic configurations of computational environments make third-party reproduction difficult, and legal and ethical issues can stand in the way of access to materials required to reproduce results. Understanding that these
predicaments are often too arduous to navigate or are simply insurmountable given technical and legal constraints, policy issuers often grant exemptions in the face of these predicaments. Expectations for policy compliance without exemption will require that technical support to resolve computational errors caused by hardware and/or software configurations that cannot be duplicated, and protocols for obtaining permission to access sensitive or proprietary data will need to be followed as data providers prescribe. **Provisions for contending with technical obstacles and other technicalities that limit data sharing and access are necessary for journal-based data policy implementation in order to resolve issues that make compliance difficult or impossible despite researchers’ best efforts.**

5.2 Study Implications

Study of the AJPS verification policy implementation and outcomes offered a useful model that illustrates how such a policy compels researchers to engage in reproducible research behaviors, and how the outcomes of the AJPS policy implementation, in particular, may inform the development of interventions that may increase researchers’ ability to meet the requirements of the policy with fewer difficulties. With strict policies, clear guidance, and accommodations for dealing with technical obstacles, compliance may be more straightforward for researchers, who likely will make the activities required to meet policy standards for reproducibility part of their normative research practice.

Results of this add to the body of literature that seeks to explain why reproducible research practices have not become part of normative research practice despite general consensus on its importance to the scientific enterprise. While studies have reported scientists’ willingness and intention to share their data, the evidence from studies that investigated the availability of datasets underlying published research results show that scientists have not implemented the
practices necessary to ensure the reproducibility of their research. The literature tells us much about what affects these practices, yet little is known about how these factors facilitate or interfere with the processes required to construct a research compendium that contain the artifacts and documentation needed to reproduce their published results. The dissertation analyzed these processes to further explain why researchers fail to meet the reproducibility standard for their own research.

Such explanations based on the particulars of processes can inform the development of interventions that support journal-based policy compliance. Considering the cost of data curation and code review for the purpose of verifying computational reproducibility, it is important that researchers fulfill policy requirements without the need for reviewers to repeat the verification workflow in instances of reproduction failure. Issuing clear, rigorous policies while providing the mechanisms that researchers need to produce high-quality research compendium may help reduce the cost of implementing rigorous journal-based verification policies, thus making it feasible for more journals to adopt such policies.

Finally, the study is part of an effort by librarians, archivists, and other research support professionals to support and promote data management best practices among scientists. This community of information professionals must be responsive to researchers’ needs as they develop support services. The effectiveness of these services depends on a clearer understanding of the specific mechanisms that facilitate reproducible research practices, which the dissertation offers.

5.4 Study Limitations

Despite the application of study validity measures, the dissertation did have some limitations that should be acknowledged and addressed.
First, the study focused on the behaviors of a specific population acting in a specific situational and disciplinary context to help explain a phenomenon experienced in the broader scientific community. The small, purposive sample of participants precludes the generalizability of results to a larger population. Moreover, the specificity of the population and disciplinary context extends to the specificity of the operationalization of computational reproducibility as it is defined by the requirements of the AJPS policy. Failure to meet some of these requirements is not always an indication of non-reproducibility. For example, the AJPS policy requires that authors include formal data citations for data obtained from extant sources. In the interest of research transparency, this information is important for understanding the analytical workflow starting from the point of data origin. However, a missing citation would not prevent an independent researcher from computationally reproducing the results, but it is an error that will trigger the resubmission process just the same as a script that fails to execute properly and produce expected outputs. The study’s focus on a narrow population acting in response to a particular operationalization of reproducibility. Thus, study results cannot claim direct application to researchers across all disciplinary domains, whose research cultures and practices may differ with respect to research reproducibility. But where expectations for reproducible research exist and where computational research occurs, such as in other social science domains that use the same or similar research tools and methodologies, study results still may prove useful.

Because of the specificity in the study and the particular focus on researchers’ processes for assembling and sharing their research compendia, the study intentionally does not consider some of the potential individual and organizational determinants of data sharing behavior investigated in previous studies. The professional rank of the corresponding author, the number
of co-authors who contributed to compendium preparation, the complexity of the analysis reported in the manuscript, and other variables may be associated with certain types of errors or higher numbers of resubmissions. Moreover, what the study concludes about reproducible research behaviors is based on the outcomes of the implementation of the AJPS data policy. As some study participants noted, publication in AJPS a top-tier journal in the political science field can have meaningful positive impact on scholars’ academic career. Whether or not the conclusion of this study would hold true for a policy issued by a mid-tier journal, and therefore support the possibility that journal-based data policies can move the needle toward adoption of reproducible research practices, remains unknown.

Finally, a critical component of the study was the author interview, which asked participants to recall past experiences. Selecting participants who had most recently published articles in AJPS reduced the amount of time between submitting their research compendium for review and the interview session. This should have made it more likely that participants’ recall more accurately reflected their actual experiences. Despite this, it is still the case that participants’ descriptions were not an exact reflection of their lived experiences given the fallibility of memory. Many of the participants, however, did refer to their own verification reports to remind them of specific issues they encountered during the verification process. In addition, some details about the nature of verification errors provided by participants were corroborated by information about the errors in their verification reports as part of the study validation strategy.

While these limitations constrain the interpretation of study results, they also call for future research that can confirm and expand on the conclusions of the dissertation.
5.4 Future Study

There are a number of areas for future study that may be pursued based on the results and conclusions of this dissertation research. Considering the study limitations described in the previous section regarding the specificity of the population of interest and context, intentional singular focus on actual behavioral controls, and shortcomings of the qualitative interviews, future investigations of researchers’ reproducible research practices may investigate other factors to determine whether actual behavioral control impacts performance of reproducible research behaviors in ways consistent with findings from the current study drawn from researchers in political science subject to a journal-based data policy.

Efforts to promote reproducible research practices are taking place within a much greater landscape of academic career pressures, competing data management and sharing mandates, disciplinary norms and emerging scientific practices, scholarly publication processes, and other concerns that influence researchers’ understanding, acceptance, and adoption of reproducible research practices. Future investigations may consider policies issued by mid-tier journals and that vary in rigor and standards of reproducibility to determine if these variances elicit different behavior. A future study may also look to other disciplinary domains to determine whether actual behavioral control impacts performance of reproducible research behaviors in ways consistent with findings from the current study drawn from researchers in political science. If differences appear, it would be worthwhile to establish why this is the case. Such an investigation may provide additional insights and nuances that further explain the mechanisms that promote adoption of reproducible research practices.

In addition to organizational factors, future studies may consider individual factors to determine their effect on actual behavioral control. As the results of the content analysis
showed, errors found in authors’ research compendia accounted for just a quarter of the variance in the number of resubmissions. What accounted for the remaining 75% of variance remains unknown. It is worth considering individual factors such as professional rank, publication record, prior data policy experience, and complexity of analysis as potential factors associated with the number of resubmissions. These additional factors can offer a more complete picture that explains scientists’ failure or success in producing reproducible research compendia.

Interpretation of the results from the current research generated new theoretical insights. Future study into researchers’ reproducible research practices may explore these theoretical insights as modal salient beliefs. As such, these salient beliefs can be used to formulate quantitative survey instrument questions. In doing so, the new theoretical insights I proposed can be tested on a larger population of researchers to investigate the applicability of these beliefs and whether or not they support the interpretation of my findings. Quantitative measurement of the correlation between successful production of reproducible research compendia and the presence of ultimatum, articulation, and accommodation in journal-based data policies can accomplish this research goal.

Finally, evidence from the study showed that previous experience with journal-based data policies inform researchers’ understanding of the policy and approach to fulfilling policy requirements. This, alongside researchers’ reports of adopting reproducible research practices (e.g., cleaner code, comprehensive documentation), suggests that education is likely a key factor in whether or not researchers succeed in producing high-quality reproducible research agenda. A future investigation might evaluate an intervention that offers data management education to researchers to determine its effectiveness in promoting journal-based data policy compliance and adoption of reproducible research practices.
APPENDIX A. Verification Report Sample

DATA REPLICATION VERIFICATION REPORT

<table>
<thead>
<tr>
<th>MANUSCRIPT NUMBER</th>
<th>[Redacted]</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARTICLE TITLE</td>
<td>[Redacted]</td>
</tr>
<tr>
<td>CORRESPONDING AUTHOR</td>
<td>[Redacted]</td>
</tr>
<tr>
<td>DATaverse DOI</td>
<td>[Redacted]</td>
</tr>
<tr>
<td>OPEN DATA</td>
<td>True</td>
</tr>
</tbody>
</table>

INITIAL SUBMISSION

| EDITOR SUBMISSION DATE | 8/7/2021       |
| VERIFICATION DATE     | 9/9/2021       |
| DATA CURATION RESULT  | Materials Incomplete |

DATA CURATION NOTES
- The figure PDFs do not need to be included in the Dataverse record as they will be in the published manuscript. CODEBOOK MISSING. You must provide a CODEBOOK in PDF that describes all variables in the analysis dataset with complete values and value labels (including missing or NA values). If the analysis data are constructed from original source data, please provide full data citations for where you collected these data.

REPLICATION RESULT       Success with Modification

REPLICATION NOTES
- All results in the manuscript replicated completely; however, there remain a few coding modifications to be made:
  - Require users to set their working directory at the beginning of the files. Don't provide a personal file path as an example; simply include `setwd()` and a comment to remind users to set the directory where the data and the helpers file are stored.
  - Provide install.packages() calls inside the scripts as well, not just in the ReadMe. These can be commented out with a reminder to users to run them as necessary.
  - reshape2 is also a required package.
  - Given that you call on the helpers file within all of the figures/tables scripts, I recommend providing a note in the ReadMe for users specifying this and describing the helper file.

Please note: Tables and figures appearing in appendices were verified only for successful execution of analysis code and not for accuracy of results.

RESUBMISSION 1

| EDITOR SUBMISSION DATE | 10/14/2021 |
| VERIFICATION DATE     | 11/8/2021   |
| DATA CURATION RESULT  | Minor Issues |
APPENDIX B. Interview Guide

1. To start, please tell me your thoughts about the AJPS verification policy.
   a. Did the policy make you consider submitting your manuscript elsewhere?

2. What were your initial expectations about the verification process?

3. Please walk me through your process of assembling and submitting your replication package for verification.
   a. What challenges, if any, did you encounter during the process?
   b. What, if anything, made the process less challenging?
   c. What kind of support did you receive during that process?
   d. How long did the process take?

4. What would have made it easier for you to assemble and submit your replication package?
   a. What kind of support would have been helpful during that process?

5. In what other instances prior to this one did you go through any of these processes?
   a. Tell me about those instances.
   b. How did that experience affect your process for fulfilling the AJPS data verification policy?

6. What was the outcome of the initial verification of the materials you submitted?
   a. What was your understanding about the issues the verifiers had about your replication package?
   b. What strategy did you use to address the issues? What was the result of that strategy?

7. How did you proceed with preparing and submitting your replication package after receiving the initial verification report?
   a. How was this process different from the initial package submission process?

8. If you could go back, what would you do differently to meet the AJPS policy requirements?

9. How has the AJPS verification policy experience affected how you deal with your research materials now?

10. What else should I know about your experience with the AJPS verification process?
APPENDIX C. Study Invitation

Dear Dr. [AUTHOR NAME],

My name is Thu-Mai Christian, and I am a PhD student who is investigating the processes researchers use to prepare data and other research materials underlying published results for sharing in a repository. Also a practicing data archivist working at the Odum Institute, I am interested in discovering how repository staff and other data support professionals can most effectively support researchers in their efforts to comply with journal-based verification policies.

Because you have published data in the American Journal of Political Science (AJPS) Dataverse, I am especially interested in your experience with the AJPS verification policy. My hope is that the study will inform the development of tools, resources, and policies that will help minimize the burden of such policies on researchers.

Your participation in the study will consist of an interview via Zoom at your convenience, with the possibility of a follow-up interview. The initial interview will include questions about your process for preparing and submitting files to comply with the AJPS verification policy. The follow-up interview, if deemed appropriate, will clarify your responses during the initial interview. The interviews will take approximately 30-45 minutes to complete, though you may end your participation at any time.

Follow the link below to take part in the research study: https://unc.az1.qualtrics.com/jfe/form/SV_07m4stSaSs0FoJE

The Institutional Review Board (IRB) at the University of North Carolina at Chapel Hill has approved this study. If you have questions or concerns about your rights as a research participant, please contact the IRB at 919-966-3113 or IRB_Subjects@unc.edu. Please reference study 22-1936. You may also contact my faculty advisor, Dr. Helen Tibbo, at tibbo@ils.unc.edu.

Sincerely,
Thu-Mai Christian
APPENDIX D. Informed Consent

University of North Carolina at Chapel Hill Research Information Sheet  
IRB Study #22-1936  
Principal Investigator: Thu-Mai Christian

The purpose of the research study is to investigate the processes used by authors to prepare and share research materials associated with published research findings. Results of the study will inform the development of interventions designed to reduce the burden of journal-based data policies on research.

You are being asked to participate in this research study because you have been identified as having authored an article in the American Journal of Political Science, which has a policy in place that requires authors to submit their data and code as part of a verification process.

Your participation in this research study is voluntary. If you agree to participate, you will be asked to take part in an interview, which will include questions about your experience in fulfilling the AJPS verification policy requirements. If necessary to clarify interview responses, you will be asked for a follow-up interview. The initial and follow-up interviews will each take approximately 30-45 minutes to complete. You can skip over any question you do not want to answer, and you may end your participation at any time.

Interviews will be conducted over the Zoom online conference platform, which offers end-to-end encryption to protect the security of users. Zoom interview sessions will be recorded only for the purpose of transcribing the interview. Interview recordings will be deleted immediately after the transcription is complete. Personally identifiable information will be omitted from the transcript. You will not be individually identified in any study reports or papers released.

This research study is being conducted independently of any journal publisher and has no effect on editorial decisions regarding manuscripts you may have under review presently or in the future. The study is also being conducted in partial fulfillment of degree requirements of the School of Information and Library Science at the University of North Carolina at Chapel Hill.

If you have any questions about this research, please contact me at tlchristian@unc.edu or 919 6983. The Institutional Review Board (IRB) at the University of North Carolina at Chapel Hill has 923 approved this study. If you have questions or concerns about your rights as a research participant, please contact the IRB at 9199663113 or IRB_Subjects@unc.edu. Please reference study #22-1936. You may also contact my faculty advisor, Dr. Helen Tibbo, at tibbo@ils.unc.edu.
REFERENCES


American Geophysical Union. (2015). Earth and space science data should be credited, preserved, open, and accessible as an integral responsibility of scientists, data stewards, and sponsoring institutions. https://www.agu.org/-/media/Files/AGU-Data-Position-Statement-Final-


Society for Information Science and Technology, 63(6), 1059–1078. https://doi.org/10.1002/asi.22634


Enke, N., Thessen, A., Bach, K., Bendix, J., Seeger, B., & Gemeinholzer, B. (2012). The user’s view on biodiversity data sharing—Investigating facts of acceptance and requirements to realize a sustainable use of research data—. Ecological Informatics, 11, 25–33. https://doi.org/10.1016/j.ecoinf.2012.03.004


foundation-expands-creative-commons-licensing-for-all-grant-funded-projects/


Survey methodology (2nd ed.). Wiley & Sons.


Stodden, V. (2013). Resolving irreproducibility in empirical and computational research. *IMS*


U.S. Department of Transportation (DOT). (2015). *Plan to increase public access to the results of*
federally-funded scientific research results.


Wallis, J. C., Rolando, E., & Borgman, C. L. (2013). If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. PLoS ONE, 8(7), e67332. https://doi.org/10.1371/journal.pone.0067332


