Transparency for Text-Based Sources: From Principles to Practice

Nikhar Gaikwad
nikhar.gaikwad@columbia.edu
Assistant Professor
Columbia University
Department of Political Science
420 W 118 St, Mail Code 3320
New York, NY 10027

Veronica Herrera
vherrera@luskin.ucla.edu
Associate Professor
University of California, Los Angeles
Luskin School of Public Affairs
337 Charles E Young Dr E
Los Angeles CA 90095

January 2022

*Acknowledgements. We thank Josh Hooper, Willow Wei, Aaron Christensen, Aura Gonzalez, Jenny Xiao for excellent research assistance. We are grateful to Rob Mickey for comments and collaboration, and for his contributions to the ideas in the section on interpretivism. This paper draws on discussions with Alan Jacobs, Tim Buthe, Kimberly Morgan and participants from the Qualitative Transparency Deliberations forum hosted by the Qualitative and Multi-Method Section of the American Political Science Association.

Transparency for Text-Based Sources: From Principles to Practice

Abstract

Debates about transparency standards in social science research often lack specificity, mischaracterize the status quo, or stress the value of replication. These debates frequently talk past each other and provide limited practical guidance for qualitative and multi-methods research. Focusing on text-based sources, this paper provides a review of qualitative research that identifies deficiencies in transparency practices, and advances a five-point framework for improving transparency premised on better specification of sources' location, production, selection, analysis, and access. We next draw on a multi-year deliberative forum on qualitative transparency to identify researchers' concerns about changing the status quo. We then showcase illustrative examples of enhanced transparency and conclude with recommendations for how to improve transparency practices for text-based sources. We argue that greater research transparency yields numerous benefits, including facilitating scholarly exchange, improving graduate training, and aiding knowledge cumulation. Rather than advancing replication, which may be undesirable for various qualitative research traditions, new transparency technologies are promising because they allow authors to more easily provide additional context, present complexity, and unpack relevant contradictions about politics.

Keywords: transparency, qualitative methods, text-based sources, archives, documents

1. Introduction

The transparency revolution has swept across the social sciences. Within political science, qualitative researchers have been divided about the value of increasing forms of transparency for qualitative data derived from interviews, focus groups, and participant observations. These discussions often talk past one another and do not systematically analyze current practices. Furthermore, these debates have not acknowledged that the costs and benefits of transparency initiatives within qualitative research differ across data types. The result has been a mischaracterization of the status quo, a failure to provide a blueprint for what transparency practices are feasible, and a lack of specificity about the costs and benefits of changing standards for research openness. In this article, we take the case of text-based sources, a data type used by scholars across the social sciences, to show that current transparency practices are insufficient, and we provide actionable guidelines for how to improve them.

Discussions about increasing transparency, or research openness, in political science began when Gary King called for the discipline to be held to the standard of replicability (1995, 444). Various initiatives followed, including the American Political Science Association (APSA)'s guidelines for the adoption of Data Access and Research Transparency (DA-RT) (APSA 2012, 9–10), new transparency technologies and institutions (Moravcsik 2014, 50–52; Kapiszewski and Karcher 2021, 473–76), and the adoption of the Journal Editors' Transparency Statements (JETS) (2015) by at least twenty-seven political science journals.

Qualitative researchers are divided over the merits of the transparency movement.[1] Some have been supportive and experimented with new developing transparency technologies (e.g. Saunders 2014, 696–97; Mayka 2021; Herrera 2015; Myrick 2021; Siewert 2021; Herrera 2017). Proponents have argued that greater transparency improves research evualation and assessment, and the research process itself; bridges diverse research communities; and facilitates knowledge building (Elman, Kapiszewski, and Lupia 2018, 39–41; Jacobs et al. 2021, 177–79).

Other qualitative researchers have pushed back against the transparency movement—questioning whether replicability can be applied to qualitative research and even the concept of transparency itself (Jacobs et al. 2021, 179–82). They have raised concerns about transparency initiatives that generate ethical issues (Monroe 2018, 143–45; Tripp 2018, 730–35) and that reflect a narrow understanding of political science scholarship, especially for qualitative-interpretive traditions (Isaac 2015, 277; Schwartz-Shea and Yanow 2016, 6–8).

While recent discussions of transparency in political science have centered on dissuading scholarly misconduct or augmenting research's replicability, this article argues that research transparency can provide more tools for evaluating claims and produce better qualitative research. Strategies for strengthening transparency come with costs—some trivial, others more substantial—and scholars often disagree about whether these costs outweigh the benefits. This article evaluates the costs and benefits of enhanced transparency, analyzes scholarly debates, and provides a practical blueprint to enhance transparency in text-based research.

We advance a five-point framework for improving transparency practices for text-based sources, premised on more explicitly specifying source location, production, selection, analysis,

[1] For an overview, see Jacobs et al. (2021, 172–74) and Elman et al. (2018, 41–43).

and access. Next, in order to explain current practices—and researchers' hesitation in changing them—we review 1,120 articles in leading political science journals. We then analyze online commentary from political scientists who participated in the Qualitative Transparency Deliberations (QTD), an online forum organized by APSA's Qualitative and Multi-Method Section from 2016 to 2018. Using posts as a primary source, we classify a set of common concerns scholars raised about research openness. These concerns are then juxtaposed against three illustrative article examples that showcase different qualitative research traditions implementing transparency in practice. We conclude by recommending a series of norms and practices for augmenting transparency for text-based sources.

2. A Framework for Evaluating Transparency in Text-Based Sources

2.1 Defining Text-Based Sources

Qualitative research depends, in part, on analyzing text-based sources. Text-based sources can include documents from government archives, records from political parties or social movements, correspondence, speeches, diaries, court rulings, media transcriptions, and secondary sources. Textual evidence may also incorporate multi-media sources such as photographs, videos, and websites. Because text-based sources are inanimate and not created by researchers, they are less prone to respondent bias (e.g. acquiescence or social desirability bias) and researcher bias (e.g., moderator or wording bias) than more interactive qualitative data collection methods such as interviews (Kapiszewski, MacLean, and Read 2015, 151–60). However, the act of selecting and interpreting text-based sources for descriptive or inferential claims can introduce other types of collection and analysis bias.

2.2 Existing Approaches for Text-Based Sources from Other Disciplines

We now turn to related disciplines to examine best practices concerning text-based sources. In history, preliminary source analysis often begins with understanding source type. For example, a source may be a relic (a physical specimen) or testimony (oral or written report). A source may have been produced intentionally (to serve as an official record) or unintentionally. Furthermore, written sources are often categorized as being either narrative (chronicles or tracts of opinion), diplomatic/juridical (documenting a legal situation), or social (products of recordkeeping by bureaucratic agencies) (Howell and Prevenier 2001, 17–28).

Marc Bloch notes that "the struggle with documents" is what defines the professional historian (1954, 86). The challenge begins with text selection. There is rarely one authoritative source, and historians must adjudicate between sources since "the majority of sources are in some ways inaccurate, incomplete and tainted by prejudice and self-interest" (Tosh 2002, 98). Political scientists encounter these challenges too, since transcripts capturing interactions between dominant and subaltern groups are imbued in power disparities where "it is frequently in the interest of both parties to tacitly conspire in misrepresentation" (Scott 1990, 2). To determine a source's authoritativeness, historians consider a document's genealogy—its genesis, its originality, and the author's trustworthiness (Howell and Prevenier 2001, 61–68). They draw inferences from a text by determining "how, when and why it came into being" (Tosh 2002, 86).

Understanding the production of sources also involves interrogating how archives are curated—whose agenda is reflected and what is absent. Archivists and state officials are responsible for how records are categorized, described, and made accessible. Context matters. Amid political uncertainty or violence, records are often lost or destroyed. In post-authoritarian countries, archival gatekeepers may fear uncovering truths about the past and decide to mislead

or intimidate researchers (Tesar 2015, 105–9). State archives may be inscrutable because of post-war gutting, financial constraints or self censorship (Daly 2017, 314–16); some public figures do not keep records (Saunders 2014, 693). When working with historical personal records, researchers may face ethical dilemmas about whether to publish findings based on sensitive materials (Tesar 2015, 110–13).

Researchers may also face ethical issues at the archives, such as sharing research on endangered languages that may offend speech communities being studied (Innes 2010, 199–202) or privacy concerns when working with psychiatric records (Taube and Burkhardt 1997, 61–63). In countries marked by economic and political instability, the best documentary evidence may be found outside of state archives (Daly 2017, 312). Informal archives can be invaluable evidentiary sources, but accessing informal archives requires the researcher to locate them and gain the owner's trust (Auerbach 2018, 345–46). We use these valuable interdisciplinary insights to inform our understanding of how to work with text-based sources.

2.3 Transparency Principles for Text-Based Sources in Political Science Research

Scholars engaged in debates about qualitative research transparency often talk past one another because there remains considerable disagreement about what research transparency means. Although the dominant typology emerging from DA-RT initiatives champions "data access, analytic transparency and production transparency" (APSA 2012, 9–10), there is little consensus on how these goals should apply to qualitative work or the overall legitimacy of the DA-RT or JETS initiatives.

We propose five transparency-enhancing principles for text-based sources. These involve increasing specificity about 1) where sources are located, 2) how sources were produced, 3) why

the researcher chose the source, 4) how the source provides evidence for the scholar's claim, and 5) how to access the source material. The overarching goal of transparency is to help others evaluate a researcher's key claims. These standards will not apply to every statement made in a manuscript, but rather to key analytical, descriptive, or causal claims (see also Jacobs et al 2021, 21).[2]

Source Location. In identifying a source's location, authors should provide enough information that readers can locate a source themselves. If a document is privately held by the author or not publicly available, this should be noted. Qualitive researchers typically situate evidentiary claims by citing and sometimes quoting a particular source, but often do not provide enough information to allow readers to actually find it. This problem is pronounced in the case of secondary source citations with missing page numbers.

Source Production. All text-based sources are produced in contexts outside of a researcher's control. These conditions present unique challenges for scholars seeking to evaluate a source's evidentiary value. When was the source created? Who was involved in its production? What were the contextual factors around its use? Was the source created by a state-sponsored organization, a media outlet with a particular ideological orientation, a paid consultant, or a political dissident? Answers to these questions can fundamentally alter scholars' interpretations. Production-related information allows readers to evaluate a source's evidentiary value through

[2] The primary responsibility of deciding which claims are key or marginal lies with authors, who can make decisions by leveraging their expertise over the subject matter. Additional clarification can be sought by reviewers during the publication process.

the broader context in which it was formed (see also Elman, Kapiszewski, and Lupia 2018, 33–34).

Source Selection. As Scott notes, sources that document the "open interaction between subordinates and those who dominate" portray very different accounts than those that occur "offstage," (1990, 2–5); thus, reliance on one type of source would provide different insights than reliance on another. Qualitative researchers working in positivist traditions are often warned against engaging in selection bias (Moravcsik 2014a, 49; Lee 2015, 2; Thies 2002, 355). Sources may be imbued with a variety of biases, which can vary depending on the nature of the research project. With these issues in mind, authors can help readers evaluate the quality and applicability of evidence by explicitly discussing why they selected specific sources to support particular evidentiary claims. Why was one source privileged over another? Is it because it provides more detail, is authored by someone with more knowledge, or simply because it is the only option available? Answers to these questions will help readers evaluate the credibility of sources and authors' use of them.

Source Analysis. Another transparency measure for text-based sources is providing information on how a source supports an author's claims—otherwise known as "analytic transparency" (Moravcsik 2014, 48–49). As Elman et al. note, "the goal of analytic transparency is to help others understand how we know what we claim to know" (2018, 34). This approach helps readers assess how the author is drawing inferences from a source or mix of sources.[3]

[3] Some readers may lack expertise in evaluating how a source supports an author's claims. Nevertheless, an author's explicit discussion of a source's analytical value clarifies her preferred interpretation and provides a baseline against which other interpretations can be evaluated.

Political scientists' record on analytic transparency is mixed. For example, some scholars are more likely than others to include discussion of a source's analytical value in a "meaty footnote," although journals' varying word-count limits and differing subfield norms lead to inconsistency across the discipline. Yet source analysis is fundamental for research transparency because it allows readers to better understand how a source supports an author's claims and why a particular document holds evidentiary value.

Source Access. The final transparency-enhancing measure involves sharing an excerpt of a source or the entire source itself. Under APSA guidelines, this technique is referred to as "data access" (APSA 2012, 9–11). There are several dimensions of research explicitness that do not involve making full or partial sources publicly available and thus greater transparency can be achieved even without data source dissemination. However, we emphasize that sharing source excerpts can help readers gauge a source's authorial intent and meaning.

2.4 Considering Interpretivism and Research Openness

Scholars in the interpretivist tradition, which centers the use of text-based sources on "meaning-making," have views that differ from the ones we present. Interpretivists and positivist qualitative researchers "increasingly do not travel under the same philosophical umbrella when it comes to…knowability of their subjects of inquiry" (Yanow and Schwartz-Shea 2015, xvii). Yet, text-based sources are central to many types of interpretivist research (e.g. Hansen 2006; Lynch 1999; Tidy 2017), and scholars rely on diverse sources such as films, postage stamps, and political cartoons as "communicators of meaning" (Yanow and Schwartz-Shea 2015, 155). Some interpretivists argue that increased research transparency can lead to greater engagement with methodological positivists, a stronger basis for interdisciplinary work, and a better understanding

of how researcher positionality impacts accessing sources, and generating and analyzing data (Yanow and Schwartz-Shea 2015, xv).

While interpretivists and positivists have different ideas about research openness, we think most would agree that—bracketing confidentiality concerns—clearly stating a source's location and the circumstances of its production are important goals. These two elements are crucial for scholars concerned with intertextuality, where texts represent a lived experience (Yanow and Schwartz-Shea 2015, 156). Source location and production help establish the context in which texts—and their interpretations—are "coproduced in and through field-based interactions rather than as objectified, free-standing entities available ('given') for 'collection' divorced from their field setting" (Yanow and Schwartz-Shea 2015, xix).

There may be less consensus between positivists and interpretivists regarding the remaining components of text-based research transparency, such as source selection. The representativeness of a text-based source, for example, would not necessarily factor into an interpretivist's decision to employ it, since providing evidence of a causal claim is not the goal of interpretivist research. Additionally, interpretivists do not view underlying texts as data until particular texts are brought into the research process; texts are devoid of meaning before a scholar's schema converts a text-as-source into a source with meaning (Yanow and Schwartz-Shea 2015, xxi). Still, interpretivists view source selection as an important component of the research process—they carefully consider the principles underlying the selection process (Hansen 2006, 73–78). Even though interpretivists and positivists have different goals, enhanced transparency in source selection might be beneficial for certain types of interpretivist work.

Analytic transparency, meanwhile, is already an established part of the interpretivist process. For interpretivists, analysis "commences when one begins to conceive of a research

project, to frame one's research question, read others' writings on the subject, and design one's study," and is not simply the "penultimate step in the research process" (Yanow and Schwartz-Shea 2015, 158). Discourse on how a source supports a given perspective or interpretation is at the core of every argument; analytic transparency is achieved through the act of interpretation itself. Thus, interpretivists who provide detailed source analysis within the body of an article might not need to include further information on analytic transparency in footnotes or appendices. These spaces may instead engage other goals such as discussing researcher positionality.

Finally, in terms of source access, interpretivists may reject the idea that different readers could come to the same conclusion by accessing the same original texts. Indeed, an interpretivist's data is "processed, not 'raw,' data—'cooked' and filtered through the initial researcher's interpretive schema"; this understanding "renders problematic the creation of databases of interpretive data for other researchers to use" (Yanow and Schwartz-Shea 2015, xxi). This does not mean that interpretivists would necessarily oppose sharing sources when possible; however, sharing sources might not indicate acceptance of expectations of a "common norm" or of the JETS transparency mandate. The sharing of texts can serve different purposes for different epistemological communities.

3. How and Why do Political Scientists Practice Transparency for Text-Based Sources?

3.1 The Status Quo

To examine the status quo in transparency practices within political science, we reviewed five leading journals during six years of publication: *American Political Science Review* (APSR), *American Journal of Political Science* (AJPS), *World Politics* (WP), *Perspectives on Politic*s

(PoP), and *Security Studies* (SS). We initially reviewed APSR, AJPS, WP, and PoP; subsequently, we selected SS precisely because it has historically been receptive to qualitative work. Publication bias may limit the range of printed scholarship; APSR and AJPS tend to publish fewer articles using qualitative methods, whereas WP, PoP, and particularly SS publish a larger amount of qualitative research (Teele and Thelen 2017, 440–41). Yet these five journals represent prestigious outlets for new work; if we are going to see extensive transparency practices in political science journals, it would likely be in these publications.

We reviewed every article in these five journals published every other year from 2008 to 2018 (a total of 1,120 articles). We then selected articles that used qualitative methods for analysis, excluding articles that used solely quantitative methods or that were mixed quantitative-qualitative methods with qualitative methods playing a very minor role. We then selected empirical articles that used text-based sources as the foundation for their claims, totalling 160 empirical articles that substantially used both qualitative methods and text-based sources; that less than 15 percent of articles surveyed made it into our sample is indicative of the low rate of qualitative research published in top political science journals.

[Table 1 around here]

Through this analysis, we identified a wide range of text-based sources used by political scientists conducting qualitative research. Table 1 shows that researchers routinely use archival material, autobiographies, court documents, government archives, pamphlets, NGO reports, newspaper articles, and secondary sources as a basis for empirical evidence. Materials employed less frequently include church documents, religious iconography, film, novels, company reports, speeches, and protest signs. The frequency with which different types of text-based sources appear across the journals varies. During the six years in our sample, SS published qualitative

research articles that used text-based sources in 218 instances, whereas the count for WP was 126, PoP was 69, APSR was 31, and AJPS was 8. Taken together, these 160 articles used text-based sources a total of 452 times.

How explicit are researchers about their use of text-based sources? For each of the five forms of transparency that we coded, an article could receive either a 1 or a 0. Table 2 summarizes our coding for transparency practices across our five dimensions, excluding articles that only relied on secondary sources; Appendix A and C provide more information.

[Table 2 around here]

When coding for source location, we focused on findability—how easily others could locate the source (if publicly available) or understand where the author had found it. Here we only coded articles whose primary use of text-based sources were not secondary sources. For example, did newspaper articles provide full titles, dates, and working html links? Did the location information include the location of the archive and additional identifying information (box number, etc.)? In our sample, 78 articles (49 percent) provided some information about where sources were located.

We see that the remaining forms of transparency were even more infrequent in our sample. Only 11 percent of articles provided information about how their sources were produced, 9 percent explained how sources were selected, 12 percent provided analysis of how sources supported claims being made, and 14 percent provided access to partial or full selections of sources. While it is not surprising that few articles provided full data access, other forms of transparency (such as providing identifying information from archives or indicating why particular sources were selected) were routinely missing.

To further analyze source location practices, we studied norms regarding the inclusion of page numbers for secondary sources (see Appendix A, fn. 3 for details about our classification of secondary sources). Table 3 confirms that omitting page numbers for secondary source citations has become the status quo. Our analysis shows that of the 20,894 total citations of scholarly sources found in the 160 articles we surveyed, only 43 percent provided page numbers for in-text citations or notes. The average masks significant heterogeneity across journals, however. In SS, which we selected anticipating stronger transparency norms in qualitative research, 67 percent of citations contained page numbers for in-text citations or notes; excluding SS, in the other four journals, only 22 percent of citations contained page numbers. In addition, 39 percent of those citations with page numbers were in cases of directly quoted text; this proportion rises to 51 percent when SS is excluded from the analysis. These findings suggest that the norm in political science has become to not use page numbers unless citing a direct quotation. That said, we find evidence that journals with stronger qualitative commitments, such as SS, have pointed a way forward by publishing more articles employing enhanced source location transparency standards (see Table 3 and Appendix A footnote 3).

[Table 3 around here]

Our review suggests that there are currently real deficiencies in qualitative researchers' methods of citing text-based sources. Studies have shown the ubiquity of "reference rot" in APSR (Gertler and Bullock 2017, 167). We frequently encountered broken URLs and missing source material on researchers' personal webpages across the journals we surveyed. Moreover, our sample presented few patterns or norms of how transparency was carried out when it did occur.

3.2 Researchers' Concerns about Changing Practices

To understand how political scientists have articulated concerns about greater transparency in research utilizing text-based sources, we draw on the Qualitative Transparency Deliberations (QTD) online discussion forum (see Jacobs et al. 2021, 171–76).[4] The QTD boards were a space for political scientists to discuss what they understood to be the costs and benefits of augmenting transparency practices in qualitative research. Participants were asked to comment on the advantages and disadvantages of increasing transparency standards for text-based sources from their current vantage point. QTD boards were broadly representative of most "typical" qualitative researchers, few of whom were active proponents or early critics of DA-RT. Participants' concerns thus provide a valuable "temperature check" on how scholars are grappling with the practical and ethical issues surrounding changing transparency norms. Concerns raised in the QTD forum allowed us to establish a clear set of research questions that we take up in the final section of this paper.

1. How to navigate copyright law? QTD discussion suggested that issues surrounding copyright infringement were unclear. Gelbman noted that "creating an expectation that source materials will be digitized opens a whole can of worms with respect to permissions and copyright."[5] If sharing entire documents becomes equated with greater transparency, how would scholars navigate this norm if they are unable to legally or ethically share entire sources? While a transparency appendix could include small excerpts (up to 150 words) from a copyrighted work containing the most relevant information, the question of sharing entire copyrighted works

[4] We examined 808 QTD blog posts for this analysis—see Appendix B for full citations.

[5] Shamira Gelbman, December 1, 2016.

worried some participants. Particular types of text-based data may face even stricter copyright restrictions. Harkness reports that in order to access archival British colonial maps, she "had to agree not to give public access to the digitized maps according to crown copyright law."[6] How would scholars, who are not trained in copyright law, determine who holds a copyright for something like photographs published in newspapers, or when a copyright expires? Researchers working internationally encounter contexts where copyright rules may be opaque, changing, or nonexistent.

2. What about archival rules and restrictions? Some archives—especially in developing countries—limit the number of pages researchers can scan or ban any reproduction of their materials. Even if an archive permits some forms of reproduction, making source materials available may be onerous and expensive, particularly for researchers with fewer resources. QTD forum participants noted that graduate students and junior faculty may struggle to scan and process large quantities of sources. Harkness reports from her experience working in African archives that "obtaining photocopies is usually possible..., although there is often a bit of an expensive racket around doing so. Those costs cannot always be well-anticipated up front..."[7] Archives might also disapprove of broad dissemination of their records. Gelbman was concerned that a requirement to share textual data, in addition to creating work for the researcher, could make archives less cooperative. She notes that "[It] can make it harder for researchers to gain access in the first place if [archives] come to believe that facsimiles of documents will be shared

[6] Kristen Harkness, April 27, 2016.

[7] Ibid.

in the peer review and publication process."[8] Gelbman also noted that "some archive[s] also change their policies, sometimes very suddenly and without maintaining record of the old rules."[9] Would researchers be able to keep up with these changes?

3. What about privacy concerns? While ethical issues may be less common in text-based research than in human-subjects research, researchers using textual evidence may also face ethical dilemmas. Even if archives permit researchers to copy documents, it could be a violation of trust with the archive to publicize its documents. Hymans writes that the owner of a private archive he worked with "did not want anyone else to see the papers, and in fact he did not want anyone else to know that he had the papers."[10] Concerns could also arise regarding the privacy of individuals described in personal communications, such as diaries or letters. Thurston notes that she uses "letters from private citizens to advocacy groups in [her] research. This raises questions about whether and how to protect their identities when it comes to citation."[11] For some research projects, efforts to de-identify documents may not sufficiently protect research subjects.

4. Can single documents be separated from their broader context? Some QTD posters worried that certain research transparency requirements could place an undue burden on qualitative researchers in cases where providing context for isolated notes requires sharing extensive background information closely linked to the research site. Several commenters were concerned that transparency measures would require separating documents from the broader

[8] Shamira Gelbman, December 1, 2016.

[9] Ibid.

[10] Jacques Hymans, December 13, 2016.

[11] Chloe Thurston, May 17, 2016.

context that only the researcher understands. One guest poster said that in their archival work, "It would be impossible for a second researcher to understand the significance of document #10 without reading documents #1-9, and #11-50."[12] If transparency always requires sharing reproduced documents in their entirety (a point we negate and take up in the final section of this article), how might authors discuss their consultation of adjacent documents for someone who had not consulted those same documents? This challenge is especially acute for interpretivist, immersive, and ethnographic scholars. Keck writes, "I am perfectly comfortable with the idea that someone who does not speak the languages I speak, does not know the history I know, and does not have the kinds of social and intellectual networks I have ...would not be able from my notes or appropriately archivable interviews to come to the same conclusions I have come to."[13]

5. What about right of first use? Another hesitation in immediately sharing original data stems from a desire to protect one's "right of first use." QTD contributors have suggested that journals requiring data sharing should also guarantee an embargo period where newly-collected data is reserved for the collector's exclusive use. Capoccia writes that "the question of embargoes for original data seems to deserve more attention from journals and research transparency advocates." Otherwise, researchers will be incentivized to use "off-the-shelf" data instead of collecting their own original sources.[14] How much lag time should be implemented when sharing data and how can such restrictions be practically implemented?

---

[12] Guest, December 9, 2016.

[13] Margaret Keck, April 8, 2016.

[14] Giovanni Capoccia, May 19, 2016.

6. How much time does transparency take? An additional cost of making qualitative research more transparent is time. Researchers may encounter a tradeoff between preparing a transparency appendix and working on other projects. Mansbridge remarks: "The work of producing a TRAX [transparency appendix] is undoubtedly good for the researcher and for the reader. But how good, compared to starting work on another important subject? We have limited lives and very limited research time."[15] Some commenters feared that adding time-consuming transparency requirements to qualitative textual research could disincentivize researchers from pursuing it. Greitens observes that "the uncertainty and increased transaction costs around qualitative research seem already to be leading many [graduate students] to conclude that the attempt is simply not worth the risk."[16] Poteete noted that "[A TRAX] would certainly slow the time to publication, potentially significantly, and that is an important cost. Slower time to publication implies fewer publications when on the job market or up for review, tenure, or promotion. Will expectations shift to account for changing practices?"[17]

7. How will transparency appendices be viewed in journal and promotion reviews? Would creating a detailed transparency appendix make one's work vulnerable to reviewers? Handlin noted, "A TRAX is not really like a quantitative appendix in this regard, which tends to present a bunch of information and additional statistical results that are carefully selected by the author to be relatively bullet proof. Presenting big sections of text from your sources in a TRAX

[15] Jane Mansbridge, October 26, 2016.

[16] Sheena Greitens, April 20, 2016.

[17] Amy Poteete, January 1, 2017.

will always open you up to potential criticism from others about the interpretation of the text."[18] Would reviewers nitpick and use the material against authors to sink a paper? Others noted that transparency appendices for qualitative research would likely be longer than quantitative article appendices; would reviewers be expected to wade through this material? Finally, how would transparency requirements affect junior faculty under pressure to publish quickly before tenure reviews?

The QTD boards broached important questions about the ethical, legal, and practical issues related to increasing transparency for qualitative research based on text-based sources, some of which are easier to negotiate than others. We take up these issues in the final recommendations section by engaging with established literature and reviewing new digital technologies that can mitigate some QTD concerns.

4. Illustrative Examples of Transparency for Text-Based Sources

We now draw on three examples of political science scholarship from across the epistemological spectrum to showcase how scholars have advanced transparency in text-based sources along the five dimensions that we previously identified. These examples illustrate how authors using diverse text-based sources and encompassing variant research goals (e.g. theory-building, identifying causal mechanisms, discourse analysis and interpretive method, and qualitative replication) have already incorporated transparency practices in their qualitative research; Appendix D provides additional examples.

[18] Sam Handlin, October 17, 2016.

Example #1: Theorizing International Legitimacy with Archival Sources

Blending a process tracing and social-constructionist approach, Musgrave and Nexon (2018) sets out to explain why states routinely invest in expensive endeavors that do not appear to yield military or economic gain. The authors argue that when states face legitimacy concerns, they carve out authority in symbolic spheres to project international leadership. This argument is advanced through analysis of diverse text-based sources—including governmental archive records, declassified intelligence documents, presidential statements, and US cabinet minutes during the Cold War. The authors also consult specialist tracts drawn from archaeological, documentary, and other primary sources from the Ming Dynasty in the early-fourteenth century.

Source Location. When introducing archival materials, the authors provide specific location information down to the last identifier, facilitating "findability" and external assessment. For example, to tie US investments in its Apollo space program to a perceived challenge to national competence after the USSR launched its Sputnik satellite, the article draws on contemporaneous viewpoints of American officials. Musgrave and Nexon provide detailed identifier information, for example: "US Information Agency, Office of Research and Analysis, 'Impact of US and Soviet Space Programs on World Opinion,' 7 July 1959, US President's Committee on Information Activities Abroad (Sprague Committee) Records, 1959–1961, Box 6, A83-10, Dwight D. Eisenhower Library, Abilene, Kansas" (Musgrave and Nexon 2018, 610, fn 107).

Source Production. In a data overview to the digital transparency appendix (or ATI, discussed in the next section) accompanying this article, Musgrave (2018) provides considerable detail regarding source production related to text-based evidence used to study state intentions in the

Ming era. Most available primary sources used in the article (e.g., "Veritable Records of the Ming Dynasty," a set of official records compiled by scholar-officials after deaths of emperors) are themselves secondary source accounts. Many primary sources had also been lost, both unintentionally (in the midst of dynastic successions) and intentionally (due to specific bureaucratic sabotage in the later Ming era). The authors therefore decided to base their analysis on secondary sources. They eschewed "standard popularizations" and instead consulted a vast body of specialist tracts (e.g., Edward Dreyer's 2007 tome, *Zheng He: China and the Oceans in the Early Ming Dynasty, 1405–1433*) that examined archaeological, documentary, and primary records. Their "capacious selection" of secondary sources permitted them to "better survey disputes over interpretations of the voyages' meaning and impact" (Musgrave 2018, data overview). Contextualizing the universe of sources relied upon allows readers to evaluate sources' evidentiary value and gain insights into excluded texts.

Source Selection. How did the authors select sources to evaluate their claim that the Ming naval expeditions to the Indian Ocean ("treasure-fleet voyages") sometimes displayed force when seeking to purchase loyalty for the emperor from local potentates? To support the claim regarding the use of coercion, the authors introduce a quote, "the Treasure-ships were intended not only to dazzle foreign peoples with their wealth and majesty but to overawe potential opposition with their might and firepower" (Finlay 1991, 12, quoted in Musgrave and Nexon 2018, 608). The supporting transparency appendix discusses the authors' decision to privilege this source, stating that a different scholar (Needham 1972, 489) provides an alternate account of the voyages, depicting them in noncoercive terms. Yet, the authors argue that Needham had a "Sinophilic outlook," which might have led him to provide an overly peaceful interpretation of

the voyages and ignore facts presented in Findley 1991 regarding their militaristic nature (Musgrave 2018, annotation 20).

Source Analysis. In the Cold War case, Musgrave and Nexon's argument hinges on the claim that US-Soviet space competition was a matter of prestige, leading to large American investments in space technologies that had no overt military or economic rationales. As part of the evidence offered to support this contention, they contrast the "romantic language of [President Kennedy's] public speeches about space" with transcripts of tape recordings of his private conversations with NASA officials, in which he states, "Everything we do ought really to be tied to getting on the Moon ahead of the Russians… [W]e ought to be clear, otherwise we shouldn't be spending this kind of money because I'm not that interested in space" (Musgrave and Nexon 2018, 616–17). The authors analyze this evidence—attaching greater weight to because it stems from a tape recording of a private conversation—to show that space exploration was motivated by prestige, consistent with the article's theoretical argument.

Source Access. To substantiate their claim that an obsession with scientific prestige quickly permeated the US establishment and motivated subsequent space investments, the authors introduce a declassified CIA document, "A Comparison of US and USSR Capabilities in Space," written in January 1960 following the launch of Sputnik. An excerpt is presented as a figure in the article and the entire document is made available in a transparency appendix. The evidentiary value of the document is strong; it rates the US and USSR on several specific dimensions, presenting what the authors argue is a "formal statement setting out how scientific

capital was exchanged into prestige" (Musgrave 2018, annotation 30; Musgrave and Nexon 2018, 610–11).

Example #2 Critiquing Dominant War Narratives with Multi-Media Sources

For an example of transparency from a different epistemic community, consider the interpretivist scholarship of Tidy (2017). Does our understanding of state-sponsored violence during wartime change when we interpret killings through the lens of ordinary citizens? Tidy's article sets out to challenge dominant narratives of Western warfare, which are created and written by those in power to advance self-serving goals and which obscure the perspectives of subordinate actors. The author draws on video footage, photographs, written narratives, media accounts, and testimony related to the killing of civilians by American forces in Iraq in 2007 to show that wartime killings can take on a political vocabulary of either "collateral damage" or "collateral murder" depending on whether they are interpreted from the perspectives of those "from above" or "from below."

Source Location. The preponderance of evidence Tidy consulted is from WikiLeaks, which dedicated a website to host videos related to an American Apache helicopter's air-to-ground attack in New Baghdad during 2007, along with press coverage, still images, transcripts, and other materials surrounding the attack. In an accompanying transparency appendix, the author provides references to WikiLeaks' online repository, along with specific links to embedded videos, stills, photographic evidence, transcripts, timelines, and additional resources (e.g., U.S. military rules of engagement; news commentary related to the event; and photos and information

about civilians killed in the attack). Links to archived copies of defunct WikiLeaks websites are also shared (Tidy 2018, annotation 1).

Source Production. Tidy carefully interrogates sources' production in order to support the article's argument that broader power structures frame popular representations of contemporary warfare. For example, the author situates videos of the killings taken from the Apache helicopter (the "view from above") within a set of dominant technologies and actors (e.g., drones and bombers) that have become the hegemonic social representation of modern warfare in the U.S. (Tidy 2017, 102). This source is contrasted with the "view from below," captured in photographs taken by a ground-level Reuters photojournalist, Namir Noor-Eldeen, until the very moment he was killed in the attack. Here, the perspective documented by a subaltern actor both physically and symbolically subordinate to the Apache crew is selected to offer a glimpse into the lives of civilians who are at a permanent disadvantage in wartime settings (Tidy 2017, 105).

Source Selection. To document the "view from below," the author selects photographs captured by Noor-Eldeen through his long-lens camera, which is significant because the Apache crew mistook the camera for an enemy missile; the source "becomes the literal and metaphorical visual mode through which the war experience of the commonly elided receivers of military violence are written into the narrative" (Tidy 2017, 105). This source is selected because it creates a Rashomon effect when juxtaposed with the "view from above," and with a third perspective, "the view from on the ground" (captured by testimony and materials from a ground-level U.S. soldier who witnessed the killings); each presents a different perspective on the same military attack, but Noor-Eldeen's visuals have a strikingly different encounter with violence.

Source Analysis. The author leverages the text and visual elements in these different sources as the basis of the article's discursive analysis. Consider the video from the Apache helicopter. As the author argues, state-sanctioned depictions of war are typically orchestrated as short clips demonstrating the mastery, precision, and rationality of military combatants; instead, this 39-minute video footage documents the Apache crew tracking targets, "interpreting an often-ambiguous feed of images," and attacking vehicles that were revealed to have children (Tidy 2017, 104). It is important to understand the weight of this source, which was classified and never intended to be public. The evidence is damning, and Tidy uses it to make a compelling critical analysis with the very tools and weapons of those who she is criticizing.

Source Access. Tidy consulted sources that were readily accessible in the public domain and provides links to them. Even though the sources are public, the article fastidiously documents pertinent information; source materials were qualitatively annotated in order to make underlying evidentiary materials easier to locate, and a transparency appendix was provided to show how specific sections of the article rely on distinct source elements (Tidy 2018).

Example #3 Qualitative Replication with Historical Documents

Our third example focuses on qualitative replication: Kreuzer (2010) revisits underlying historical sources used in a quantitative study of the origins of proportional representation systems to replicate its findings. Kreuzer argues that differences in how sources are selected, interpreted, and translated into numerical datasets substantively alters the conclusions of quantitative analyses.

Source Location. Kreuzer illustrates why it is important to be able to locate sources utilized in historical research. When assessing the reliability of the Labor Market Index, a key quantitative measure used in Cusack, Iversen, and Soskice (2007) ["CIS"], Kreuzer finds that the "absence of citations for 31 of the 90 cells (34.4%) in Table 1 makes it extremely difficult to replicate" CIS's codings (2010, 373). In Appendix B, Kreuzer documents how the sources referenced by CIS provide evidence for only certain constituent parts of the Labor Market Index, leaving other components unsubstantiated. The inability to locate sources fundamentally matters; Kreuzer's qualitative replication presents alternate sources that disprove CIS's findings.

Source Production. Kreuzer also disputes several of CIS's coding decisions by calling into question sources' production value. For example, Kreuzer (2010, 372) argues that CIS use Colin Crouch (1993) as their primary source for the strength of guilds, employer associations, and union centralization, yet fails to explain *how* cases were coded. In Appendix A, meanwhile, Kreuzer (2010, 385) draws on a number of alternate country-specific historical sources to question these coding choices. For example, he relies on evidence in Jost (1990, 284–86) and Ebbinghaus (1995, 73) to recode Switzerland as not having centralized unions because Swiss unions were divided by language and religion.

Source Selection. CIS's finding that economic determinants led to the adoption of proportional representation systems derives from an analysis of 18 cases. Kreuzer conducts an in-depth historical examination into each of the 18 cases, consulting political, labor and economic history sources (2010, 375). The replication from this expanded set of sources casts

considerable doubt on the casual mechanisms in CIS's argument. Different types of sources can lead to very different conclusions; more information on source selection can enhance transparency practices considerably.

Source Analysis. In order to dispute CIS's coding that France did not have widespread rural cooperatives, the author provides historical evidence regarding France's rural cooperatives. Kreuzer (2010, 372) directs the reader to Appendix A, where he cites two studies (Cleary 1989, 40-50; Mares 2003, 133-34) that confirm that France "experienced a rapid growth of agricultural associations from the 1890s onward…Their growth continued throughout the interwar period." This evidence supports Kreuzer's decision to reverse the CIS's coding.

Source Access. After examining the qualitative evidence, Kreuzer recodes 13 out of CIS's 90 coding choices (2010, 371–72). For example, CIS codes Austria as having a large skill-based export sector; in Appendix A, Kreuzer provides evidence to show that exports constituted only a small percentage of Austria's output. He provides a source excerpt from Katzenstein (1985, 138) to note that in the late nineteenth century, "Austrian producers by and large eschewed the specialization for exports typical of the other small European countries" and provides evidence from several other sources to support this claim. Taken together, Kreuzer makes a strong case for why enhanced transparency practices can advance replication and social scientific enquiry.

5. Recommendations for Researchers

The types of transparency practices showcased in our three illustrative examples are unfortunately not common in the majority of articles we surveyed previously. Instead, most

articles provided little information about their sources. We now offer recommendations aimed at improving this state of affairs. We surveyed existing literature to generate recommendations for work using text-based sources that includes the use of 1) detailed citations, 2) transparency appendices, and 3) data access, In doing so, we differentiate between practices that should be required, ones we strongly encourage, and those we view as optional. In these recommendations, we address concerns raised in the QTD boards and provide practical solutions for authors.

1. Detailed Citations Should be Just Like Real Estate: "Location, Location, Location"

Requiring detailed citations specifying the location of cited evidence so that others may easily find referenced works should become a foundational practice in qualitative political science. We brought to bear systematic data to confirm previous studies decrying imprecise citations practices (e.g. Lustick 1996, 6; Trachtenberg 2015, 13–14; Moravcsik 2010, 30) with our review that revealed low page number citation rates in top disciplinary journals. Political science journals replaced the use of discursive footnotes with parenthetical references in shifts to mirror quantitively-oriented fields (Lustick 1996, 6; Trachtenberg 2015, 14); furthermore, the APSA Style Manual is vague about when to use page numbers (APSA 2018, 39). The result has been that citations have become merely ornaments as opposed to facilitators of scholarly interchange. We cannot improve transparency practices without addressing this fundamental problem.

1A. Specify page numbers for scholarly sources (require). We propose that page numbers be required for most in-text citations for journal submissions when citing scholarly sources such as articles, books, and book chapters. For any scholarly source used within the text or notes, manuscripts should cite the precise page or page ranges where evidence is being drawn.

When referencing the main argument of a work, citing page numbers guides readers toward a summary of the main argument, but citing entire books or articles does not help others easily consult sources. We propose that journals require authors to report the percentage of their citations that have page numbers and set a benchmark that submissions have to meet (e.g. 75%) or otherwise be desk rejected.

1B. Specify location of primary sources—where a source resides and where evidence resides within a source (require when ethically and legally possible). Other text-based sources, such as those identified in Table 1, are what historians call "primary sources" or sources that serve as an original foundation of information about a topic under study, often from firsthand witnesses. We urge scholars to specify the location of such primary sources whenever ethically and legally possible. This includes information from all publicly available archives, policy reports, etc. We strongly encourage qualitative researchers to cite like historians, who reference archival material down to the last identifier so that others can "pull the box from the shelf" (Trachtenberg 2015, 14). Detailed location citations would include 1) citing the location where the source resides, and 2) specifying the location within the source itself where evidence is being pulled to support claims. The default practice wherever possible should be to cite the location within the source itself where an evidentiary claim is being pulled to the last available identifier (e.g., with a time stamp for an audiovisual, not just the title and date). If the source originates from the private archive of someone who wants to remain anonymous, the author should say so and provide de-identified information where ethically and legally possible. Online-only sources are likely to have different enumeration, but authors should simply include any available identifying location, such as a chapter number or section heading (see also APSA 2018, 45). We urge editors to specify the practice of location "findability" as "best practice" in their submission

protocol and for reviewers of manuscripts using text-based sources to ask authors to make necessary revisions during R&Rs.

2. Transparency Appendices: New Technologies in a Changing World

As political science journals have adopted scientific notation practices and tight length limits, the practice of providing extended source annotation in footnotes has waned. We argue that manuscripts should provide information about their text-based sources, including how they were selected and produced, and most importantly, how they provide evidence for key claims that support their argument. We discuss two practices that scholars can adopt to improve transparency standards for text-based sources: 1) a text-based source overview and 2) source-based annotation.

2A. Text-Based Source Overview (strongly encourage). Most journal articles based on text-based sources provide only a cursory methods statement within the text itself. Our review revealed that political science manuscripts rarely treat text-based sources as a form of data that requires extended explanation. We recommend that authors provide a methods narrative in an appendix that includes a statement about the types of text-based sources used and relevant information about how they were produced and selected. Here authors can speak in a holistic way about their data generation, because data citations cannot capture the large amount of material that was used to draw inferences and generate conclusions. Moreover, some research traditions rely on evidence gathered over a long period of time, or depend on deep background knowledge and inferential paths that defy step-by-step descriptions (Elman, Kapiszewski, and Lupia 2018, 34). Source overviews allow authors to specify how particular documents were interpreted in light of other documents consulted.

Authors should decide the length of text-based source overviews—from a paragraph to a few pages—and withhold information as necessary to respect any legal and ethical constraints about their sources. For example, an author can specify 1) the overall universe of information relevant to the central research question, 2) how she decided what to consult, record, and what quotes to select for analysis, 3) what relevant data (e.g., boxes in archives or secondary sources) were not consulted or were omitted and the potential impact on the analysis and 4) how key sources were produced and came into the author's possession (Elman et al. 2017, 7; Elman, Kapiszewski, and Lupia 2018, 33).[19] Methods appendices are common for top university press books; we argue that such a standard should also be adopted for journal articles that publish research based on text-based sources. Importantly, this type of appendix would not count against an article's word limit.

All research, and particularly qualitative research, relies on interpretation—different people may come to different conclusions when looking at the same source. Replication is not necessarily feasible, or even desirable, for many types of qualitative research (Jacobs et al. 2021, 171–85). Yet providing more information about text-based sources is not really about replication, but rather helping others understand the evidentiary basis of an argument and facilitating scholarly debate. The form and content of such an overview would reflect the diversity of the many epistemological communities within our discipline.

[19] The Qualitative Data Repository (QDR) encourages the use of an ATI Data Overview, which is an approximately 1,000-word length description of the research process. See https://qdr.syr.edu/ati/ati-instructions.

2B. Source-Based Annotation: From Meaty Footnotes to Annotation for Transparent Inquiry (ATI) (encourage). Individual annotations linked to referenced evidence is the hallmark of excellent qualitative scholarship. This practice is ubiquitous in history, legal scholarship, and social sciences scholarship published in historically oriented journals like *Studies in American Political Development (SAPD)*. Discursive annotations allow authors to explain why the cited reference undergirds key claims. Annotations can also provide a space to explain source selection, production, and relevant bias within the source.

We recommend source-based annotation in political science be strengthened by adopting two practices. First, annotations in the form of footnotes or endnotes should not count against the journal word limit, as this practice discourages transparency and scholarly exchange, particularly for research that relies on text-based sources. Journals should adopt the practice of "transparency footnoting," where word limits exclude footnotes or endnotes (e.g. see *Politics & Society*). Alternatively, journals with longer word limits, such as *International Security* (20,000 words) or *SAPD* (no official limit) provide authors with more space to augment transparency practices in footnotes and endnotes. We agree with Gerring and Cojocaru (2020, 98) that length limits are arbitrary and counterproductive.

Second, we encourage the use of ATI, a digital annotation infrastructure hosted by the Qualitative Data Repository. The ATI is automatically available for all Cambridge University Press journals, but is compatible with all journals. Popular press articles, newspapers, and social media now use hyperlinks to draw readers to digital information that provides supporting evidence for claims, illustrates examples, or facilitates further reading. As scholars we should not hesitate to do the same. Based on Moravcsik's pioneering active citation work (2014; 2010), an ATI is based on "open annotation" for generating and sharing digital annotation across the web

that allows any digitally published manuscript to be annotated (Kapiszewski and Karcher 2021, 473–74). ATIs allow authors to share excerpts from sources and provide more detailed information relating to how specific passages within a journal article support key claims.

The ATI design provides a good blueprint for how we might think about annotating sources regardless of whether we use ATI, a different future technology, or a simply a meaty footnote. When deciding what to annotate, the ATI architects suggest considering 1) the centrality of the evidence-based claims, 2) the importance of the data source, 3) whether a claim is contested or controversial and 4) whether a source is contested or controversial (Elman et al. 2017, 11–12). ATI architects recommend providing a full citation, an analytic note discussing how the data was generated and analyzed to support the claim, a source excerpt (typically 100-150 words), and the data source itself if can be shared legally and ethically (Elman et al. 2017, 3). ATIs are flexible: authors can choose to provide annotations for as little or as much of their article as they like. They can provide excerpts without annotations, annotations without excerpts, and links to full data sources as appropriate. ATI architects urge annotations for only a subset of passages, and explicitly state that it "…is unnecessary, potentially counter-productive, and almost certainly time-consuming" to annotate all passages that involve descriptive or causal inference (Elman et al. 2017, 11).

Although ATIs are relatively new, users have identified numerous benefits. Annotations allow authors to engage directly with inevitable contradictions that emerge in data by providing more space to adjudicate between conflicting pieces of evidence (Mayka 2021, 480; Myrick 2021, 493; Siewert 2021, 489). In addition, ATIs can serve multiple qualitative research communities. For Qualitative Comparative Analysis, ATIs can help corroborate coding decisions by connecting original sources to their final assessment and make interpretations of evidence

more transparent (Siewert 2021, 488–89). For multi-method qualitative and quantitative research, annotations can supplement technical material to make the paper's quantitative analysis more accessible (Myrick 2021, 493). For process tracing studies, ATIs allow researchers to showcase the logic of inductive or deductive analysis of causal process observations or applications of process tracing tests (Mayka 2021, 480). Authors can use ATIs to share participants' own words to better illustrate the lived experience of one's interlocutors and how they engage in meaning-making (Mayka 2021, 480). ATIs even allow authors to share digital content such as maps, posters, songs or videos.

ATI users have also noted opportunities for improving its uptake, including considerations of when and what to annotate and how to reduce the primary cost of annotating: time. The time it takes to compile transparency appendices was a key concern on the QTD boards. In our experience (Herrera 2015), creating an ATI does take time, but integrating ATIs into the article drafting process can greatly reduce the time required for annotations (Herrera 2017; Mayka 2021, 481). Also, more proactive messaging from journals about how to integrate ATI into the peer review process (a QTD concern) is needed (Myrick 2021, 495).

QTD posters sometimes worried that qualitative transparency appendices would need to be very lengthy and recreate the entire research process. ATI users suggest just the opposite, emphasizing the utility of concise and tailored annotations. Bombarding readers with too many annotations is counterproductive and burdensome for both readers and authors (Myrick 2021, 494; Siewert 2021, 490). Authors should fit annotations tightly to the top priorities of the manuscript. Mayka suggests potentially self-imposing a maximum number of annotations to safeguard against wading "aimlessly in [a] sea of qualitative data" (2021, 481). Siewert

repeatedly interrogated specific information to determine whether it was "essential" before annotating (2021, 490).

ATIs users will have to decide what to annotate in the digital ATI format and what to leave for a footnote in the journal text; the two should complement each other rather than overlap. ATIs are most useful for annotating key claims and sources, as they allow an author to discuss a source's location, production, and selection process in detail and outside the scope of an article world count. ATIs should be viewed as a new approach that allows researchers to maintain control over their research agenda, epistemological commitments, and the ethical and legal considerations specific to each individual research project.

3. Data Access: Excerpts vs. Archiving Data access—and the thorny question of data sharing—was the most contentious component of the QTD debate, but in our estimation, the most misunderstood. Proponents of qualitative data sharing have consistently maintained that data sharing should be optional and adhere to legal, ethical, and logistical constraints (e.g. Kapiszewski and Karcher 2021, 473, 477; Elman et al. 2017, 1, 2, 4; Elman, Kapiszewski, and Lupia 2018, 42). We distinguish between two options that were often conflated in the QTD boards, 1) Sharing excerpts of text-based sources, which we encourage and 2) sharing an entire text-based source, which we view as optional.

Our section on ATIs discussed sharing data excerpts; here we clarify how it compares to sharing entire text-based sources. Proponents of qualitative data sharing are frequently referring to the sharing of excerpts (typically 100-150 words), which can be accomplished with ATI technology, meaty footnotes, or a supplementary appendix. Sharing excerpts is little more than sharing an extended quotation of a source. QTD participants expressed concerns about navigating copyright law. Written excerpts of fewer than 150 words are subject to fair use

copyright law, which allows for the reproduction of short excerpts for scholarly purposes (Karcher, Kirilova, and Weber 2016, 295; APSA 2018, 7).[20] If copyright rules of non-text based sources such as photographs or maps are unattainable or face restrictions outside of US law, authors can simply decide to not post the source.

QTD boards also broached the subject of privacy concerns when sharing text-based sources. Sharing excerpts of a text-based source will not be controversial for the majority of pre-existing sources (e.g., constitutions or organizational charters). Scholars working with sensitive documents or privacy concerns can choose to share de-identified information or simply not share. While de-identified data and analytic utility may at times be at odds (e.g. Kapiszewski and Karcher 2020, 209, note 10), researchers always retain authority over what excerpts to share and how to present them to others as de-identified sources.

Alternatively, researchers may choose to archive entire research projects, although this option will likely only appeal to those interested in facilitating future data use by others. While qualitative data archiving is relatively new in political science, there are examples of existing archived qualitative projects in other fields (Mannheimer et al. 2019, 645–46). Several repositories exist for archiving social science data that offer data curation and preservation, but the Qualitative Data Repository at Syracuse University is an ideal choice because it was specifically designed to accommodate the heterogeneity of qualitative data source types and is curated by political scientists (Kapiszewski and Karcher 2020, 205–2012; Mannheimer et al. 2019, 647).

Data repositories can help guide authors on how to legally and ethically share sources and even help researchers write data management plans. Institutional repositories can council

[20] See Appendix E.

researchers on how to de-identify data to preserve anonymity, work with university IRB procedures and standards for informed consent, create differential or restricted access (so sensitive material can be shared only by "request"), and preserve the right of "first use"[21] (Kapiszewski and Karcher 2020, 205; Mannheimer et al. 2019, 649–53). If a particular de-identified document would be rendered analytically useless or would insufficiently address legal and ethical concerns, it can simply not be shared. As for the QTD board concern regarding archival policies, authors who are bound by archival rules that do not permit sharing would simply not share. If scholars have an organized data management plan, time spent depositing files will likely be manageable (Saunders 2014, 695; Kapiszewski and Karcher 2020, 201–2). Sharing data through a dedicated repository as opposed to author websites makes the data more accessible over time, and repositories are also better than journals at archiving large quantities of documents[22] (Kapiszewski and Karcher 2020, 205–6; Mannheimer et al. 2019, 647).

While data archiving may sometimes incur fees and does take time, the benefits are multiple. Data archiving can boost authors visibility when their data is reused and cited (Kapiszewski and Karcher 2020, 217–18), aid in pedagogical purposes (Kapiszewski and Karcher 2020, 198), promote organized workflow, facilitate analysis and writing (Mannheimer et al. 2019, 650), and allow others to more easily engage with one's scholarship. Archiving is a promising, optional choice for some research projects.

---

[21] APSA recommends a "right of first use" embargo period of one year or periods specified by the journal, press publishing claims, or relevant funding agency (APSA 2012, 10).

[22] See Herrera and Mayka (2020) for an example of online archived text-based sources via a journal's interface (supplementary materials includes a methods narrative and 10 pdfs).

6. Conclusion

Qualitative researchers frequently engage in careful, painstaking work with difficult-to-access texts in demanding research settings. They do so in different languages and in politically volatile contexts. Yet publishing trends have shrunk the space available for scholars to explain which sources were consulted, why they were selected, and how they support key claims. As a result, citations have become symbolic placeholders rather than facilitators of scholarly exchange, and scholars are confused about how and whether to modify existing practices.

Our paper contributes a practical element to this complex debate by focusing on how transparency can be augmented in research that relies on text-based sources.[23] Expanding Moravcsik's original formulation (2014, 48–49), we specify five types of text-based source transparency dimensions (location, production, selection, analysis, and access). We provide a first-of-its-kind assessment of existing practices, reviewing 1,120 articles published between 2008-2018, and find that less than 15 percent of articles provided information about source production, selection, analysis, and access. Most sources fail to provide information about where sources are located. Page number use is extremely low at 22 percent for *APSR*, *AJPS*, *World Politics* and *Perspectives*, and of these citations, half are for quoted text. Our review provides evidence that there is considerable room for improvement.

We use the QTD deliberations, a multiyear forum for online discussion about transparency practices for qualitative research, to identify researchers' key concerns—such as navigating copyright law, archival rules, privacy concerns, right of first use, time burdens, and hiring and promotion impacts. We spell out recommendations for augmenting transparency in

[23] Appendix F provides additional sources for transparency issues in qualitative research.

the use of text-based sources organized around 1) detailed citations, 2) transparency appendices, and 3) data access, using illustrative examples and established literature to showcase how others have navigated challenges. Our recommendations begin with those we see as most urgent, mandating the use of page numbers, to those we view as optional, such as data archiving.

We also outline the benefits of new technologies, such as ATIs, that allow scholars from diverse epistemological traditions to annotate and share their text-based sources via excerpts or full sources where legally and ethically feasible. ATIs will not solve challenges inherent to qualitative research, such as divergent evidence, noncomparable data types, "the absence of evidence as evidence," and the entanglement between data collection and analysis that often makes the research process difficult. However, ATIs provide more space to wrestle with these issues, even if it is not a panacea.

Transparency technologies for qualitaitve research are unlikely to advance if departments do not count ATIs or data archiving towards hiring, tenure, and promotion. Data files could count as half or one third of a peer reviewed article, and should be showcased on CVs, findable and citable. Letter writers can highlight their value for a hiring or promotion file. We see valuing ATIs or data archives in the same light of counting a research note in a promotion file—not as a means to punish scholars who have not adopted such measures but rather provide professional recognition for those that do.

Incorporating more training about transparency technologies into qualitative research in graduate school would help scholars mitigate potential costs. Graduate qualitative methods courses can instill transparency norms in the research apparatus of early-stage scholars. Students can gain experience with the five dimensions of transparency by studying the practices highlighted in our illustrative examples or by incorporating similar practices in their own

research. Paying attention to transparency in the early stages of research projects (e.g., gathering source excerpts during data collection or drafting with potential annotations in mind) reduces time burdens. Learning transparency technologies as an early stage qualitative researcher is also likely to produce other benefits, such as enhanced training surrounding working with documentary evidence and archives, and increased proficiency in multimedia data organization and project management.

The practical guidelines we propose have implications for other types of research. Although QTD participants disagreed fervently about the benefits and costs of data sharing, our paper dispels the myth that data sharing (in particular, uploading entire documents) is a required and necessary part of research openness. Instead, greater transparency can be achieved through the use of excerpts or through other forms of transparency such as providing information about sources' location, selection, production and analysis. Other types of qualitative research, such as those based on elite interviews, process tracing, or participant observation, could develop a similar approach that deemphasizes the need for data sharing and clarifies a path forward for enhancing transparency within particular epistemological traditions. There are implications for quantitative studies as well. Quantitative scholars frequently choose among data sets that vary in quality; applying the principles of transparency in "source selection" and "source production," scholars should provide information on why they selected particular datasets and discuss potential biases associated with them.

Enhanced transparency measures such as those we outline here are beneficial and frequently possible, and can be adopted in ways that retain authorial authority and epistemological commitments. New transparency technologies are promising because they allow qualitative

researchers to more easily provide more context, present complexity, and unpack relevant contradictions about politics.

References

APSA. 2012. *A Guide to Professional Ethics in Political Science*. Second. Washington D.C.: American Political Science Association Committe on Professional Ethics, Rights and Freedom.

———. 2018. *Style Manual for Political Science: Revised 2018 Edition*. Washington D.C.: American Political Science Association.

Auerbach, Adam. 2018. "Informal Archives: Historical Narratives and the Preservation of Paper in India's Urban Slums." *Studies in Comparative International Development* 53 (3): 343–64.

Bloch, Marc. 1954. *The Historian's Craft*. Manchester Univ Press.

Daly, Samuel Fury Childs. 2017. "Archival Research in Africa." *African Affairs* 116 (463): 311–20.

Elman, Colin, Diana Kapiszewski, and Arthur Lupia. 2018. "Transparent Social Inquiry: Implications for Political Science." *Annual Review of Political Science* 21 (1): 29–47.

Elman, Colin, Diana Kapiszewski, Andrew Moravcsik, and Sebastian Karcher. 2017. "A Guide to Annotation for Transparent Inquiry (ATI) Version 1.0 (Prepared for the ATI Pilot Working Group, Workshop #1)," December, 1–13.

Gerring, John, and Lee Cojocaru. 2020. "Length Limits." In *The Production of Knowledge: Enhancing Progress in Social Science*, edited by Colin Elman, John Gerring, and James Mahoney, 98–126. Cambridge: Cambridge University Press.

Gertler, Aaron L., and John G. Bullock. 2017. "Reference Rot: An Emerging Threat to Transparency in Political Science." *PS: Political Science & Politics* 50 (01): 166–71.

Hansen, Lene. 2006. *Security as Practice: Discourse Analysis and the Bosnian War*. London: Routledge.

Herrera, Veronica. 2015. "Data for: "Does Commercialization Undermine the Benefits of Decentralization for Local Services Provision? Evidence from Mexico's Urban Water and Sanitation Sector." Qualitative Data Repository. 2015. http://doi.org/10.5064/F6F769GQ.

———. 2017. "Understanding the User Experience: Lessons from an Active Citation Piloteer." *Qualitative Data Repository Blog*, October. https://qdr.syr.edu/qdr-blog/understanding-user-experience-lessons-active-citation-piloteer.

Herrera, Veronica, and Lindsay Mayka. 2020. "How Do Legal Strategies Advance Social Accountability? Evaluating Mechanisms in Colombia." *Journal of Development Studies* 56 (8): 1437–54.

Howell, Martha, and Walter Prevenier. 2001. *From Reliable Sources: An Introduction to Historical Methods*. Ithica: Cornell University Press.

Innes, Pamela. 2010. "Ethical Problems in Archival Research: Beyond Accessibility." *Language & Communication* 30 (3): 198–203.

Isaac, Jeffrey C. 2015. “For a More Public Political Science.” *Perspectives on Politics* 13 (2): 269–83.

Jacobs, Alan M., Tim Buthe, Ana M. Arjona, Leonardo R. Arriola, Eva Bellin, Andrew Bennett, Lisa Bjorkman, et al. 2021. “The Qualitative Transparency Deliberations: Insights and Implications.” *Perspectives on Politics* 19 (1): 171–208.

JETS. 2015. “Journal Editors’ Transparency Satement (JETS).” *Dartstatement.Org*.

Kapiszewski, Diana, and Sebastian Karcher. 2020. “Making Research Data Accessible.” In *The Production of Knowledge: Enhancing Progress in Social Science*, edited by Colin Elman, John Gerring, and James Mahoney, 197–220. Cambridge: Cambridge University Press.

———. 2021. “Empowering Transparency: Annotation for Transparent Inquiry (ATI).” *PS: Political Science & Politics* 54 (3): 473–78.

Kapiszewski, Diana, Lauren MacLean, and Benjamin Read. 2015. *Field Research in Political Science: Practices and Principles*. Cambridge: Cambridge University Press.

Karcher, Sebastian, Dessislava Kirilova, and Nicholas Weber. 2016. “Beyond the Matrix: Repository Services for Qualitative Data.” *IFLA Journal* 42 (4): 292–302. https://doi.org/10.1177/0340035216672870.

King, Gary. 1995. “Replication, Replication.” *PS: Political Science & Politics* 28 (03): 444–52.

Kreuzer, Marcus. 2010. “Historical Knowledge and Quantitative Analysis: The Case of the Origins of Proportional Representation.” *American Political Science Review* 104 (2): 369–92.

Lustick, Ian S. 1996. “Read My Footnotes.” *APSA-CP: Newsletter of the APSA Organized Section in Comparative Politics* 7 (1): 6,10.

Lynch, Cecelia. 1999. *Beyond Appeasement: Interpreting Interwar Peace Movements in World Politics*. Ithaca: Cornell University Press.

Mannheimer, Sara, Amy Pienta, Dessislava Kirilova, Colin Elman, and Amber Wutich. 2019. “Qualitative Data Sharing: Data Repositories and Academic Libraries as Key Partners in Addressing Challenges.” *American Behavioral Scientist* 63 (5): 643–64.

Mayka, Lindsay. 2021. “Bridging Positivist and Interpretative Approaches through Annotation for Transparent Inquiry.” *PS: Political Science & Politics* 54 (3): 479–82.

Monroe, Kristen Renwick. 2018. “The Rush to Transparency: DA-RT and the Potential Dangers for Qualitative Research.” *Perspectives on Politics* 16 (01): 141–48.

Moravcsik, Andrew. 2010. “Active Citation: A Precondition for Replicable Qualitative Research.” *PS: Political Science & Politics* 43 (01): 29–35.

———. 2014. “Transparency: The Revolution in Qualitative Research.” *PS: Political Science & Politics* 47 (01): 48–53.

Musgrave, Paul. 2018. “Data for: Defending Hierarchy from the Moon to the Indian Ocean: Symbolic Capital and Political Dominance in Early Modern China and the Cold War.” Qualitative Data Repository. 2018. https://doi.org/10.5064/F6NNOI5C.

Musgrave, Paul, and Daniel Nexon. 2018. “Defending Hierarchy from the Moon to the Indian Ocean: Symbolic Capital and Political Dominance in Early Modern China and the Cold War.” *International Organization* 72 (3): 591–626.

Myrick, Rachel. 2021. “Reflections on Using Annotation for Transparent Inquiry in Mixed-Methods Research.” *PS: Political Science & Politics* 54 (3): 492–95.

Saunders, Elizabeth N. 2014. “Transparency without Tears: A Pragmatic Approach to Transparent Security Studies Research.” *Security Studies* 23 (4): 689–98.
Schwartz-Shea, Peregrine, and Dvora Yanow. 2016. “Legitimizing Political Science or Splitting the Discipline? Reflections on DA-RT and the Policy-Making Role of a Professional Association.” *Politics & Gender* 12 (03).
Scott, James. 1990. *Domination and the Arts of Resistance: Hidden Transcripts*. New Haven: Yale University Press.
Siewert, Markus. 2021. “How Annotation for Transparent Inquiry Can Enhance Research Transparency in Qualitative Comparative Analysis.” *PS: Political Science & Politics* 54 (3): 487–91.
Taube, Daniel O., and Susan Burkhardt. 1997. “Ethical and Legal Risks Associated with Archival Research.” *Ethics & Behavior* 7 (1): 59–67.
Teele, Dawn Langan, and Kathleen Thelen. 2017. “Gender in the Journals: Publication Patterns in Political Science.” *PS: Political Science & Politics* 50 (2): 433–47.
Tesar, Marek. 2015. “Ethics and Truth in Archival Research.” *History of Education* 44 (1): 101–14.
Tidy, Joanna. 2017. “Visual Regimes and the Politics of War Experience: Rewriting War ‘from above’in WikiLeaks’‘Collateral Murder.’” *Review of International Studies* 43 (1): 95–111.
———. 2018. “Data for: Visual Regimes and the Politics of War Experience: Rewriting War ‘from above’ in WikiLeaks’ ‘Collateral Murder.’” Qualitative Data Repository. 2018. https://doi.org/10.5064/F643DB5O.
Tosh, John. 2002. *The Pursuit of History: Aims, Methods, and New Directions in the Study of Modern History*. Harlow: Longman.
Trachtenberg, Marc. 2015. “Transparency in Practice: Using Written Sources.” *Qualitative and Multi-Method Research*, 13–17.
Tripp, Aili Mari. 2018. “Transparency and Integrity in Conducting Field Research on Politics in Challenging Contexts.” *Perspectives on Politics* 16 (3): 728–38.
Yanow, Dvora, and Peregrine Schwartz-Shea. 2015. *Interpretation and Method: Empirical Research Methods and the Interpretive Turn*. New York: Routledge. Ebook.

Transparency for Text-Based Sources: From Principles to Practice

Perspectives on Politics

Tables

Table 1: Variation in Usage of Text-Based Source Types in Political Science Journals

	APSR	AJPS	WP	PoP	SS	Totals
Archival Material	5	2	3	5	34	49
Autobiographies	1	0	5	0	11	17
Blog Posts	1	0	0	0	0	1
Campaign Documents	0	0	0	1	0	1
Church Documents, Religious Iconography	0	0	2	0	0	2
Company Reports & Business Documents	0	0	2	0	0	2
Constitutions	1	0	1	0	0	2
Court Documents & Deliberative Internal Court Proceedings, and Laws	1	0	3	6	0	10
Decrees	0	0	1	0	0	1
Film & Novels	0	0	1	1	1	3
Legislation & Parliamentary Debates	0	0	3	1	0	4
Government Documents, Archives, and Pamphlets	3	2	13	9	28	55
Letters	1	0	1	0	3	5
Magazines	0	0	2	0	0	2
NGO & IGO Reports	1	0	15	6	16	38
Newspapers	5	1	23	10	43	82
Organizations Publications (e.g., Muslim Brotherhood)	1	0	1	1	3	6
Party Documents	0	0	1	0	1	2
Protest Signs	0	0	1	0	0	1
Secondary Sources*	10	3	42	27	70	152
Speeches	1	0	1	1	6	9
Truth & Reconciliation Commission Reports and Tribunal Documents	0	0	2	0	0	2
Websites	0	0	1	1	0	2
WikiLeaks	0	0	1	0	2	3
World Bank Reports & UNDP Report	0	0	1	0	0	1
Totals	31	8	126	69	218	452

Source: Authors' compilation. *Secondary sources in this table refer mostly to historical interpretations (e.g., historical treatises, military histories, etc.) See Appendix A, fn 1 "A note on secondary sources."

Table 2: Transparency Practices for Text-Based Sources ("TBS") (2008–2018)

	Total Number of Empirical Articles using TBS	Articles with Source Location Information	Articles with Source Production Information	Articles with Source Selection Information	Articles with Source Analysis Information	Articles with Source Access Information
APSR	13	3 (23%)	2 (15%)	1 (8%)	3 (23%)	2 (15%)
AJPS	3	2 (67%)	1 (33%)	0 (0%)	0 (0%)	0 (0%)
World Politics	45	23 (51%)	6 (13%)	9 (20%)	5 (11%)	4 (9%)
Perspectives on Politics	27	14 (52%)	4 (15%)	2 (7%)	1 (4%)	5 (19%)
Security Studies	72	36 (50%)	5 (7%)	3 (4%)	10 (14%)	11 (15%)
Total	160	78 (49%)	18 (11%)	15 (9%)	19 (12%)	22 (14%)

Source: Authors' compilation, see Appendix A. Cells include raw counts and percentages in parentheses.

Table 3: Use of Page Numbers in Secondary Source Citations

	No. of Scholarly Source Citations	No. of Scholarly Source Citations Using Page Numbers	No. of Quoted Text for Scholarly Source Citations	No. of Scholarly Source Citations using Page Numbers/ No. of Scholarly Source Citations	No. of Quoted Text for Scholarly Source Citations/ No. of Scholarly Source Citations using Page Numbers
APSR	1,746	490	261	28%	53%
AJPS	428	30	21	7%	70%
World Politics	5,769	1,217	555	21%	46%
Perspective on Politics	3,238	756	430	23%	57%
Security Studies	9,713	6,508	2,230	67%	34%
Total	20,894	9,001	3,497	43%	39%
Total, excluding Security Studies	11,181	2,493	1,267	22%	51%

Source: Authors' compilation, see Appendix A.

Transparency for Text-Based Sources: From Principles to Practice

Perspectives on Politics

Supplementary Materials: Appendix A-F

February 2022

Nikhar Gaikwad, Columbia University
Veronica Herrera, UCLA

Appendix A: Coding Procedures for Review of Existing Transparency Practices

We reviewed every article published in the *American Political Science Review*, the *American Journal of Political Science*, *World Politics*, *Perspectives on Politics and Security Studies*[1] published every other year over a six-year period (2008, 2010, 2012, 2014, 2016, 2018), which included a total of 1,120 articles. Of these, 160 were empirical articles using qualitative methods relying on text-based sources. We excluded articles that used solely quantitative methods or that were mixed quantitative-qualitative methods with qualitative methods playing a very minor role. We were inclusive in coding the use of qualitative methodology; for example, articles that were quantitative but also used illustrative vignettes or background cases were included. We excluded articles whose primary or sole empirical basis were interviews, participant observation or ethnography, and we also excluded articles that were not empirical, (e.g., political philosophy). We then selected empirical articles that used text-based sources as the foundation for their empirical claims. We included articles that used at least five text-based sources as units of analysis (at the unit level, e.g., a constitutional article, as opposed to the source itself, e.g., a constitution) as their empirical foundation. That is, five text-based sources "TBS" could be five passages from a constitution, even though the constitution is one source; the unit, here, is the passage, not the constitution.

We found that for the five years reviewed, 160 empirical articles used qualitative methods as well as text-based sources as a substantial component of the methodological approach and evidentiary basis of claims. The fact that less than 15% of articles surveyed made it into our sample likely reflects the publication choices of top political science journals; a similar percentage would have likely resulted had we chosen a different group of top journals. Of these 160 articles, only 28 included supplemental material or appendices of some sort.[2]

---

[1] *Regarding the inclusion of Security Studies:* We initially reviewed APSR, AJPS, WP, and PoP; subsequently, we selected SS precisely because it has historically been receptive to qualitative work and allows for submissions of up to 15,000 words, allowing comparisons with the transparency practices of the other four journals. Our selection method also does not capture transparency practices in other types of publications, such as book chapters or books, the latter of which is likely to provide more opportunities for research transparency due to higher word count limits.

Our analysis shows that of the 20,894 total citations of scholarly sources found in the 160 articles we surveyed, only 43 percent provided page numbers for in-text citations or notes. The average masks significant heterogeneity across journals, however. In SS, which we selected anticipating stronger transparency norms in qualitative research, 67 percent of citations contained page numbers for in-text citations or notes; excluding SS, in the other four journals, only 22 percent of citations contained page numbers. In addition, 39 percent of those citations with page numbers were in cases of directly quoted text; this proportion rises to 51 percent when SS is excluded from the analysis.

[2] We note that other forces may have contributed to trends in transparency documented here, such as copyeditors asking authors to delete page numbers where direct quotations are not referenced, lack of clarity for norms regarding

One person coded all of the articles after corroborating intercoder reliability for a small subset of articles with a second researcher; we did not conduct further intercoder reliability tests. We used SCOPUS as our search engine. Below we outline the article selection and coding procedures we used through two key steps. In the spirit of transparency, we also include notes from our coder as they proceeded through this research exercise.

Step 1: Article Selection Process

- First generate search for all articles for the year. Scopus: Journal Title (Source Title), year (2018, 2016, 2014, 2012, 2010, 2008). Limit search to articles only. Each article in the relevant year needs to be examined through abstract, and then opened. Determine whether it belongs in "codeable" pile. If it does, save a copy, and enter details in spreadsheet (author year). Download and also save appendices and supplementary materials (same author year, plus appendices and supplementary materials).
- To be entered into codeable count, article needs to:
    a. Use text-based sources in a significant way as evidentiary base through a qualitative analysis.
        i. Thus, primarily quantitative studies can be excluded.
        ii. Illustrative vignettes or background cases that are important parts of evidence count if they rely on text-based sources (as opposed to e.g. interviews)
    b. Be empirical (as opposed to pure theory). (Hints: Does article have a methods section? Does article propose to analyze texts or sources in systematic way for descriptive or causal argument? Is article written by a political theorist or claims to develop political theory? Articles by political theorists that include empirical material using text-based sources as a primary form of evidence were included).
    c. Use (some, specifically > 5) text-based sources as evidence for argument being made. Note that the count is at the unit level rather than the actual source itself. (e.g., 5 TBS "sources" could be five passages from a constitution, even though the constitution is one source. The unit, then is the passage, not the constitution in terms of the count used to determine codeable status). When in doubt, count the source. Examples of TBS types: Secondary sources[3] (when being used as

referencing source locations in online journals, style manuals not updating citation standards for social media sources, and other publishing pressures.

[3] *A Note on Secondary Sources:* We define secondary sources as scholarly journal articles, books and book chapters that are typically written by scholars for an academic audience. We follow the discipline of history and define secondary sources as those created by someone who did not experience first-hand or participate in the events or conditions being described or analyzed. While secondary sources can be viewed as text-based sources broadly defined, we excluded these sources from our coding analysis for source location, production, selection, analysis and access except for the coding we performed for determining page number citation percentages, where we only included secondary sources or scholarly sources such as journal articles, books and book chapters (i.e., Table 3).

However, sometimes secondary sources can be used in ways that approximate the use of primary sources in qualitative empirical research in political science. For the purposes of our coding of source types presented in Table 1, we included two types of secondary sources that were used for analysis in qualitative empirical research. These fell under three broad categories, although the first was the most common 1) books and journal articles that were *historical interpretations* (e.g. historical treatises, military histories, histories of presidential administrations, etc.), 2) *policy analysis* from a scholarly source that were descriptive pieces about a particular policy (e.g. Bush doctrine for

empirical evidence in main body of paper to support argument being made, not in literature review); Archival material; Government data; Newspaper articles; Organization's publications; NGO reports; Resolutions, Accords, Constitutions; Multimedia (e.g., video, audio clips, maps).

Step 2: Code Article along five categories:
Location, Production, Selection, Analysis, Access

- In the spreadsheet, make a note of the primary format for transparency communication: in-text, footnote/endnote, appendix, or methods statement.
- Highlight in blue the places where transparency communication is happening within the article PDF wherever relevant or feasible. Save a copy of the highlighted version.

1) Source Location.
    a. Did the primary TBS being used get discussed in terms of where it is located such that another person could find it? Code 0 or 1. Especially regarding foreign language items, was it in original language so the source could be located by others? When it is translated and the link is broken it is very difficult to find the source). This may vary across data source type, for example. Source location was searched within-text, in footnotes/endnotes, appendices, and methods statements. Additional points to consider:
        i. *Newspaper articles*: Full title, full date, html link. (Author may not be necessary if source can be found without author). Needs html or other indication to be complete (especially if you search online and can't find the source—oftentimes these links get broken).
        ii. *Archival materials*: Where in the archive, box #s, etc., and location of the archive?
        iii. (*Organizational) Resolutions, Mandates,* etc. need names, titles, location within document (e.g. pg # or para, or article), but also information about where they are physically located
        iv. *Online data sources.* Requires names, html links, dates accessed.
    b. Notes from our coder on how they managed the coding process for this category: "I would first look at the type of TBS being used. Secondary sources were straightforward to determine location. For other documents, I would look in footnotes or bibliography. URLs were the easiest way of determining location but I would also look for physical location and then see if the physical location of the text was specific enough for another researcher to find as well. So, for example, just mentioning a city, would not qualify. Physical texts would need an easily mappable location with perhaps a building, stack number, or reference number within a specific set of archives so other researchers could easily access the source without having to dig to find it."

2) Source Production
    a. Did the primary TBS being used get discussed in terms of how it was produced, by whom, etc.? Beyond just citation, did the authors give background information

foreign policy) and 3) *Foreign language scholarly analysis* on a wide range of specialized issues that political scientists drew on to help build case studies or other qualitative analysis.

as to what person or institution actually made the document? Did they discuss anything about what circumstances surrounded the making of the document, what biases or agendas were being furthered by the making of the document, or the level of trustworthiness, expertise, etc., related to the creators of the document? If archival material, was any information on the archive and how it is curated included? Code 0 or 1.

b. Notes from our coder on how they managed the coding process for this category: "Production was rarely discussed by any of the authors. I gave production a 1 if they gave useful background information on the organizations that produced the documents or if they gave information on the context of the document. An example of a 1 is Valdez 2016 on p. 19. The author gives the background context for the passage of the law before writing about the law itself. Through this contextualization, the author gives the audience a lens through which to understand the law before using it as evidence. Another example of this is Gotham 2012 who gives context on why FEMA and DHS were created before using documents produced by the two agencies. I found that if production is not mentioned before the sources are used, then it likely won't be mentioned. I checked footnotes and supplementary material to confirm this."

Valdez, Inés. 2016. "Nondomination or Practices of Freedom? French Muslim Women, Foucault, and The Full Veil Ban." *American Political Science Review* 110(1): 18–30.

Gotham, Kevin Fox. 2012. "Disaster, Inc.: Privatization and Post-Katrina Rebuilding in New Orleans." *Perspectives on Politics* 10(3): 633–646.

3) Source Selection
   a. Did the primary TBS being used get discussed in terms of how it was selected by researcher for use in this study? Does author explain what made them use this source instead of another? Does the author discuss anything about why they are using the source? Code 0 or 1.
   b. Notes from our coder on how they managed the coding process for this category: "Froio 2018 p. 701 in *Perspectives* is a good example of selection. The author gave a clear roadmap of what made them include certain sources and exclude others. After checking the supplementary materials and footnotes, I would scan the article and see if the author provided any background information on their sources. If they did, I would decide if it either fell under production or selection. For selection, the author would have to go beyond just contextualizing the source and actually describe how they arrived at using the source to make their claim. Selection was given a 1 if you could easily tell why the author chose to use the source they did over other relevant sources."

Froio, Caterina. 2018. "Race, Religion, or Culture? Framing Islam between Racism and Neo-Racism in the Online Network of the French Far Right." *Perspectives on Politics* 16(3): 696–709.

4) Source Analysis

- Note: This category was the most difficult to code, it was a more subjective category than the other four.
    a. Did the primary TBS being used get discussed in terms of how the source supports the claims being made? Did this occur consistently throughout? More than just a simple citation, does the author draw the line from the source to their claim? Do they do this consistently with most of the main text-based sources being coded in our spreadsheet? Code 0 or 1.
    b. Notes from our coder on how they managed the coding process for this category: "I looked to see if the author first explains what the original TBS was arguing, or what it was saying, before the author used it to support their argument. Did the author describe what the original source was saying and then link it to the point they were making? It was easy to eliminate an article if the author seemingly used citations but then gave no evidence or explanation to what part of the author's argument the citation is supporting. If direct quotes were used, I looked to see if they explained the quotes and directly applied them to their argument (Green 2008, *Perspectives*; Baccini & Koenig-Archibugi 2014, *World Politics*). Meyer 2016 is also an example of good analysis; the author makes a point, uses an in-text citation, and then expands upon the point or pulls in a direct quote that explained the use of the citation in the footnotes. Analysis was given a 1 if there was little room for interpretation on how the claim being made is supported by the TBS, direct quotes were the most useful in determining this category. Another key factor for receiving a 1 was consistency; if the author consistently explained with direct quotes or showed how the source used supported the claim, then they would receive a 1."
    c. As noted above, coding for analytical transparency was the most difficult of the five categories because of the many different epistemological communities and research traditions represented across the articles surveyed, and the diverse manner in which text-based sources were employed in articles. Yet we believe that it is a possible to create a decision rule that captures this category so that it is codable in a large review such as the one we undertook in this paper. The decision rule that we chose was that an article had to have had at least three (3) instances of analytic transparency connected to a text-based source in order to be coded as demonstrating analytic transparency. However, in order to create sensitivity bounds for this analysis, we randomly selected and reviewed 1/3rd of our sample in order to ascertain results from a more permissive decision rule.[4] In particular, under the more permissive decision rule, an article was coded as demonstrating analytic transparency if it had just one (1) or more instances of analytic transparency connected to a text-based source. When an article was borderline in terms of being coded for analytic transparency, we gave it a "1". We then created sensitivity bounds by studying the number of articles that were coded as demonstrating analytic transparency and by extrapolating this proportion to the entire sample of 160 articles. Under this more permissive coding, the total number

[4] We randomized the selection by starting with the first article in our sample and then selecting every fifth article across each journal in the order *APSR, AJPS, Perspectives, World Politics, Security Studies*.

of articles receiving a "1" for analytic transparency for the 160 articles would increase to 33, or 20% of the sample. That is, while our original coding resulted in 12% of articles in the entire sample being classified as demonstrating analytic transparency (see Table 2), the more permissive coding scheme resulted in 20% of articles in the entire sample being classified as demonstrating analytic transparency, which is still quite a low proportion. This permissive coding scheme is an upper bound since articles merely needed to have only one (1) instance of analytic transparency connected to a text-based source to get classified as demonstrating analytic transparency.

Baccini, Leonardo and Mathias Koenig-Archibugi. 2014. "Why do States Commit to International Labor Standards? Interdependent Ratification of Core ILO Conventions, 1948–2009." *World Politics* 66(3): 446–490.

Green, Elliott. 2008. "Understanding the Limits to Ethnic Change: Lessons from Uganda's 'Lost Counties'." *Perspectives on Politics* 6(3): 473–485.

Meyer, Brett. 2016. "Learning to Love the Government: Trade Unions and late Adoption of the Minimum Wage." *World Politics* 68(3): 538–575.

5) Source Access
    a. Did the primary TBS being used get shared or have excerpts from the source shared? For quotes from the same source type, were authors consistently using quotes that encompassed most of the point being made? Must be recurring. If URL is provided, does URL bring you directly to source (or source excerpt)? Should not have to dig at all, source should be there, in its entirety, through link provided, or in the appendix or on the author's website. Code 0 or 1.
    b. Notes from our coder on how they managed the coding process for this category: "To determine if something deserved a 1 or a 0, I would first look at the section of the paper that used qualitative evidence (see Column E in the workbook). I would take note of the citations used and then go to either the footnotes, end notes, supplementary material, or bibliography to see if there was a link to the full source or if the source was replicated. Even if there was a URL, I would then confirm that it led to the source and was still an active website and not a broken link."

Appendix B: Full Citations for the Qualitative Transparency Deliberations Board Posts

For a detailed description of the QTD background, objectives, and process, see: www.qualtd.net/page/about/ and also Jacobs et al (2021, 172–76). A complete archive of the online deliberations is available on Harvard Dataverse: https://doi.org/10.7910/DVN/SWVFV8. See Stage 2-Working Group II.1. Text based sources.

Full Citations

Kristen Harkness, "What might qualitative data access look like?," QTD Discussion Board, April 27, 2016.

Shamira Gelbman, "Documenting use of text-based or non-text-based sources," Discussion forum II.1, QTD Discussion Board, December 1, 2016.

Jane Mansbridge, "Benefits and Costs of Increasing Transparency for Text and Non Text Based Sources," Discussion forum II.1, QTD Discussion Board, October 26, 2016.

Jacques Hymans, "Benefits and Costs of Increasing Transparency for Text and Non Text Based Sources," Discussion forum II.1, QTD Discussion Board, December 13, 2016.

Chloe Thurston, "Active citation versus the meaty footnote," QTD Discussion Board, May 17, 2016.

Guest, "Documenting use of text-based or non-text-based sources," Discussion forum II.1, QTD Discussion Board, December 9, 2016.

Margaret Keck, "No place for my work in this debate," QTD Discussion Board, April 8, 2016.

Giovanni Capoccia, "Data access and 'right to first use' of newly collected quantitative data," QTD Discussion Board, May 19, 2016.

Sheena Greitens, "DA-RT: effect on graduate training" QTD Discussion Board, April 20, 2016.

Sam Handlin, "Benefits and Costs of Increasing Transparency for Text and Non Text Based Sources," Discussion forum II.1, QTD Discussion Board, October 17, 2016.

Amy Poteete, "Benefits and Costs of Increasing Transparency for Text and Non Text Based Sources," Discussion forum II.1," QTD Discussion Board, January 1, 2017.

Appendix C: Additional Results from Review of Published Articles

Table 1: Number of Published Articles using Text Based Sources (2008-2018)

	American Political Science Review	American Journal of Political Science	World Politics	Perspective on Politics	Security Studies
2018	2	1	8	9	14
2016	7	0	10	5	11
2014	1	1	4	2	13
2012	2	1	12	4	11
2010	0	0	6	4	11
2008	1	0	5	3	12
Total	13	3	45	27	72

Note: We only included articles whereby the methodology (form of analysis) used was primarily qualitative. Within this subset of articles, we excluded articles whereby the predominant evidence base was not text based sources (e.g. interviews, ethnography). For qualitative research using text-based sources, we excluded those that only used secondary sources.

Appendix D: Additional Illustrative Examples of Transparency in Practice

In the main manuscript, we provide a detailed analysis of three published articles that implement transparency practices across our five dimensions of transparency. In this section, we provide several additional examples from scholarship drawing from different subfields and corners of the discipline in both journal articles and scholarly books in order to further illustrate that many scholars are already practicing forms of transparency about text-based sources and how they do so.

First, transparency about source location helps other researchers locate data and evaluate it, expanding the scope and reach of one's research. For appropriate assessment of research based on texts, it is necessary to provide detailed information about where sources are located if they are publicly available. Providing detailed page numbers when sources are used as empirical evidence, or providing specific archival location information down to the last identifier, is an important first step toward "findability" and external assessment. For example, Kimberly

Johnson provides excellent identifier information in her archival research on African American poll-tax registration and voting when she specifies, "This article draws from the following archival sources: Luther Porter Jackson Papers, 1772-1960, Accession No. 1952-1, Special Collections and Archives, Johnston Memorial Library, Virginia State University, Petersburg, VA, hereafter referred to LPJ papers." She also references her sources down to the last possible identifier available—for example, "'New Canton Voters Association Meeting (Minutes),' May 1946, Folder 548, Box 19, LPJ" (2017, 220, fn 13, 231 fn 73).

Similarly, Sarah Goodman goes beyond citing a newspaper title and date, a practice which has become all too common. In her work on citizenship and immigration, Goodman cites her source as "Philip Johnson, 'Migrants Face New "Britishness" Test.' *Telegraph*. December 5, 2006" (Goodman 2012, 687, fn 88). By providing the newspaper article's author, the title in the original language, the newspaper name, and the exact date, it is more likely that others can find the source in the future even if original URLs are broken. In making citations, researchers should go beyond what reviewers or editors require for specifying location information for text-based sources. They should consider what subsequent readers will need to know about where publicly available sources are located in order to evaluate or build on their scholarship.

Second, specificity about a source's production allows readers to better understand the quality of the data being used and its appropriateness for a particular evidentiary claim. For example, in *Networks in Contention: The Divisive Politics of Climate Change*, Jennifer Hadden explains how she was "subscribed to the private internal listservs of Climate Action Network Europe and Climate Action Network International, as well as to the public lists of Climate Justice Alliance and Climate Justice Network. During the time period of this study, I collected more than 10,000 emails through these channels, which kept me exceptionally well informed

about the political process and the workings of these coalitions. As all these emails are confidential, I do not discuss their contents except when they exist in the public sphere or have been discussed in on-the-record interviews" (Hadden 2015, 194). By disclosing the private origins of her source, Hadden adds credence to her evidentiary claims regarding the inner workings of a climate activism organization. While this work may not be replicable due to the private nature of the source, the information given on the source's production nevertheless boosts the author's claims.

Specificity about source production can occur even for informal archives where authors may have permission to access archives but not share their contents. In his study on slum leadership and public service provision in India, Adam Auerbach photographed and digitally organized materials curated by local leaders in informal archives and described how these materials were produced and maintained. Auerbach explained how local leaders chose to preserve some materials while others were destroyed or deteriorated over time, stressing that "the decision to maintain materials is a choice that, like documentation, is non-random" (2018, 360). Auerbach described how photographs, petitions, letters from officials, party materials, and community meeting notes provided key insights, but also how they were often incomplete. Individuals were less likely to share documentary information that might cast them in a negative light, and local leaders' were more apt to keep particularly well-maintained written records of citizen petitions that documented how they as leaders engaged in problem-solving efforts (Auerbach 2018, 351, 354). Furthermore, important events were left out of the informal archival materials when they related to the Hindu minority who were socially isolated in the Muslim community (Auerbach 2018, 359–60). Auerbach's description of how his text-based sources were produced help readers

understand their evidentiary value as well as their limitations. It also demonstrates the importance of triangulating with other sources, such as interviews and ethnography.

Third, specificity about a scholar's source selection process provides insights into her methodology, the degree to which a source was "cherry-picked," and the context for the different types of biases that may present. For instance, Alisha Holland uses newspaper articles to measure attitudes about public perceptions on street vending in Lima and Bogotá in her book *Forbearance as Redistribution: The Politics of Informal Welfare in Latin America* (Holland 2017, 164–68, 334–37). In Chapter 4 and in "Appendix B: Coding Rules for Campaign Platform, Newspaper and Administrative Sources," Holland describes how she constructed a database of these articles. She specifies the date range of the search and offers several categories of news articles selected (news items, letters to editors, short notes, platforms, speeches, interviews). She then coded content based on whether or not it was sympathetic to the practice of street vending—with discussions regarding unemployment, displacement, criminalization of poverty coded as sympathetic, and discussions regarding public space, business competition, public safety risk, and crime coded as unsympathetic. Furthermore, Holland explains why she chose to code newspaper articles and not rely on public opinion polling (it didn't exist), when she used online or physical newspapers, why she chose the newspapers she chose, and her triangulation strategies with information from campaign platforms and administrative sources.

In a similar vein, Ruth Rubin identifies what archival sources were available to her and how representative they are in her book, *Building the Bloc: Intraparty Organization in the US Congress* (2017). Rubin selected organizations and their archives for study in order to cover a range of contextual variation. She argues that these archives, not the official Congressional Record, are the most appropriate source on intra-party organization because the researcher can

"follow members of Congress off the floor and behind closed doors" (Rubin 2017, 26). The book also notes source limitations, stating that, "These data are far from perfect. Members of Congress do not preserve all of their correspondence and may not be entirely forthcoming, even in confidential letters" (Rubin 2017, 26). Rubin notes the absence of data in many places, including how the "lack of formal records" among Republican Senate insurgent organizations "complicates our analysis of the lawmakers' organizational activity. In the absence of organizational records, we must rely first and foremost on individual records" (Rubin 2017, 72). However, she leverages these data limitations to support her argument, proposing that, "We can view the organization's lack of formal record keeping as a data point in and of itself" (Rubin 2017, 72). Rubin also notes how a broader evidentiary base lends credence to the cited source. For instance, she identifies the "strong archival evidence" for press coordination between House and Senate Republican insurgents, then cites a particular letter emblematic of this evidence base (Rubin 2017, 47, fn 62). The book's appendices include a summary of all archival collections she consulted, lists of all the meetings whose minutes she reviewed, and excerpts of key documents.

Fourth, specificity about how a source is interpreted and analyzed helps readers assess why the sources being provided are indeed evidence for the author's claims. For example, in his article on how converts to Protestant Christian sects influenced democracy-building in Asia, Robert Woodberry makes the claim that "Christian converts published the first privately printed Japanese-and Korean-language newspapers" (Woodberry 2012, 250, n. 17). He provides detailed reasoning in a "meaty footnote" that helps substantiate his claim by noting, alongside additional citations, that,

> The first privately printed Japanese-language newspaper was printed by Hamada Hikizō /Joseph Heco, a Protestant who had worked with missionary printers, and Kishida Ginkō, a student of the missionary Joseph Hepburn….An earlier government-printed paper was a translation of a Chinese-language missionary newspaper, minus the religious content. The

> Japanese government distributed it to a small number of high government officials as a way to monitor the outside world. It was not available to the public. The first privately printed Korean-language newspaper (the Independent) was edited by Philip Jaisohn/So Chaep'I, a Protestant teacher at a mission school. Missionaries encouraged him to publish it, provided the trained printing staff free of charge, and continued printing the paper after he fled Korea….

In providing this explanation, Woodberry strengthens the validity of his claim and signals scholarly rigor. Word count limits at various publications have made analytical transparency a more challenging task for qualitative researchers. Nevertheless, describing the analytical route through which authors link their claims to sources is critically important for augmenting transparency.

An added benefit to this transparency practice is that it establishes a trail for researchers to follow and "retrace" their steps throughout the research process, especially when data archiving is added to the mix. Producing a transparency appendix during the actual research process is also less burdensome than scrambling to create one in the last stages of publication (Saunders 2014, 696). For example, Veronica Herrera notes that in her experience, developing a transparency appendix "in the moment," (rather than retroactively) facilitates the research and writing process by providing a built-in organizational structure.[5]

Finally, sharing excerpts or full replications of text-based sources provides a plethora of benefits. Excerpts from sources can help other researchers understand why a scholar is using certain language to make key claims and how that language is linked to evidence from original sources. For example, in his book *The Rights Revolution: Lawyers, Activists, and Supreme Courts in Comparative Perspective*, Charles Epp argues that Canada underwent a major rights

[5] Herrera also noted that online data archiving can be a highly useful system for keeping files organized, especially if one's computer and backup crash, which happened to Herrera in 2015. She was able to restore her files because they were stored in a QDR transparency appendix (at the time a pilot project as an Active Citation Compilation, now an Annotation for Transparent Inquiry) (Herrera 2015).

revolution for the rights of the accused in the 1980s, and that NGOs and other organizations helped provide support for this rights revolution. He shows how leading Canadian civil liberties associations participated in key revisions to the Canadian Constitution's Charter of Rights and Freedoms' wording on search and seizure procedures, exclusion of evidence, detention and imprisonment, and the right to counsel and jury trials (Epp 1998, 188). Epp's claims are strengthened through his thirty-two line endnote with detailed excerpts of the before and after wording of provisions in the Charter regarding each of these legal rights (Epp 1998, 276–77, n. 59).

Providing a full source for a document that is in the public domain but that is hard for other researchers to find can constitute an academic public good. For example, in their study on social accountability strategies for healthcare and environmental policy in Colombia, Veronica Herrera and Lindsay Mayka (2020) create a text-based source appendix that provides the full texts of publicly available sources cited in their paper, including high court rulings, reports from the offices of the Attorney General and Ombudsman, and eighteen newspaper articles. Even though these materials are available to the public, they were gathered during field research and would not be easy for others to find without similar fieldwork. Although most researchers likely have a personal document management system for their individual workflow, organizing documents in a way that prioritizes transparency and access for other researchers will have additional benefits. Authors will be less likely to cut corners and more likely to add additional contextual information to their source notes when they anticipate how others will access this information to assess the project's contents.

Ultimately, "transparency cannot substitute for good research" (Saunders 2014, 694). However, research transparency does provide important benefits to the field—such as signaling

scholarly rigor, improved communication of evidence and findings, the development of research subfields and topics as more primary sources and data become widely available, improved qualitative research training for graduate students, and increased ease of scholarly exchange.

Appendix E: Copyright and Fair Use Resources

APSA Style Manual's Summary of Fair Use Doctrine

"The fair use doctrine, which has developed over time, is seen as one of the corner stones of free expression in the United States. Fair use limits copy right to balance the interests of copyright holders with public interest in the wider distribution and use of creative works… Fair use is determined by (1) the purpose and character of use, whether for commercial or educational purposes, (2) the nature of the copyrighted work, (3) the amount and substantialness of the portion in relation to the whole, and (4) the effect of the use on the potential market for the copyrighted work… The use of an entire literary work in its entirety is hardly ever accept-able. Use that is not fair will not be excused by paraphrasing, as it is considered disguise copying by copyright doctrine (4.89)." (APSA 2018, 7).

Stanford Libraries, "What is Fair Use?"

"In its most general sense, a fair use is any copying of copyrighted material done for a limited and "transformative" purpose, such as to comment upon, criticize, or parody a copyrighted work. Such uses can be done without permission from the copyright owner. In other words, fair use is a defense against a claim of copyright infringement. If your use qualifies as a fair use, then it would not be considered an infringement," Stim (2019, unknown page numbers). For more information, see https://fairuse.stanford.edu/overview/fair-use/what-is-fair-use/.

Websites

Fairuse.stanford.edu (Stanford Libraries, Copyright and Fair Use, Copyright FAQs; Fair Use)

Copyright.gov

Richard Stim's website: dearrichblogspot.com

Appendix Bibliography

APSA. 2018. *Style Manual for Political Science: Revised 2018 Edition*. Washington D.C.: American Political Science Association.

Auerbach, Adam. 2018. "Informal Archives: Historical Narratives and the Preservation of Paper in India's Urban Slums." *Studies in Comparative International Development* 53 (3): 343–64.

Epp, Charles. 1998. *The Rights Revolution: Lawyers, Activists, and Supreme Courts in Comparative Perspective*. Chicago: Chicago University Press.

Goodman, Sara Wallace. 2012. "Fortifying Citizenship: Policy Strategies for Civic Integration in Western Europe." *World Politics* 64 (4): 659–98.

Hadden, Jennifer. 2015. *Networks in Contention: The Divisive Politics of Climate Change*. New York, NY: Cambridge University Press.

Herrera, Veronica. 2015. "Data for: "Does Commercialization Undermine the Benefits of Decentralization for Local Services Provision? Evidence from Mexico's Urban Water and Sanitation Sector." Qualitative Data Repository. 2015. http://doi.org/10.5064/F6F769GQ.

Herrera, Veronica, and Lindsay Mayka. 2020. "How Do Legal Strategies Advance Social Accountability? Evaluating Mechanisms in Colombia." *Journal of Development Studies* 56 (8): 1437–54.

Holland, Alisha C. 2017. *Forbearance as Redistribution: The Politics of Informal Welfare in Latin America*. New York: Cambridge University Press.

Jacobs, Alan M., Tim Buthe, Ana M. Arjona, Leonardo R. Arriola, Eva Bellin, Andrew Bennett, Lisa Bjorkman, et al. 2021. "The Qualitative Transparency Deliberations: Insights and Implications." *Perspectives on Politics* 19 (1): 171–208.

Johnson, Kimberley S. 2017. "From Politics to Protest: African American Voting in Virginia in the Pre–Civil Rights Movement Era, 1940–1954." *Studies in American Political Development* 31 (2): 218–37.

Rubin, Ruth Bloch. 2017. *Building the Bloc: Intraparty Organization in the US Congress*. New York: Cambridge University Press.

Saunders, Elizabeth N. 2014. "Transparency without Tears: A Pragmatic Approach to Transparent Security Studies Research." *Security Studies* 23 (4): 689–98. https://doi.org/10.1080/09636412.2014.970405.

Stim, Richard. 2019. *Getting Permission: Using & Licensing Copyright-Protected Materials Online & Off*. 7th ed. NOLO.

Woodberry, Robert D. 2012. "The Missionary Roots of Liberal Democracy." *American Political Science Review* 106 (2): 244–74.

Appendix F: Reading Lists on Transparency Issues for Qualitative Researchers

Below we provide sources for further reading on qualitative research transparency. The list is incomplete but nevertheless a good starting point.

Further Reading from Qualitative Data Repository Affiliated Authors (found on QDR website, downloaded on 1.28.22)

Jacobs, Alan M., Diana Kapiszewski, and Sebastian Karcher. 2022. "Using Annotation for Transparent Inquiry (ATI) to Teach Qualitative Research Methods." PS: Political Science & Politics 55 (1): 216–20. https://doi.org/10.1017/S1049096521001335.

VandeVusse, Alicia, Jennifer Mueller, and Sebastian Karcher. 2021. "Qualitative Data Sharing: Participant Understanding, Motivation, and Consent." Qualitative Health Research. doi: 10.1177/10497323211054058.

Kapiszewski, Diana, and Elisabeth Jean Wood. 2021. "Ethics, Epistemology, and Openness in Research with Human Participants." Perspectives on Politics 1-17. https://doi.org/10.1017/S1537592720004703.

Demgenski, Robert, Sebastian Karcher, Dessi Kirilova, and Nic Weber. 2021. "Introducing the Qualitative Data Repository's Curation Handbook." Journal of eScience Librarianship 10 (3): 1207. https://doi.org/10.7191/jeslib.2021.1207.

Kapiszewski, Diana, and Sebastian Karcher. 2021a. "Transparency in Practice in Qualitative Research." PS: Political Science & Politics 54 (2): 285–91. https://doi.org/10.1017/S1049096520000955.

———. 2021b. "Empowering Transparency: Annotation for Transparent Inquiry (ATI)." PS: Political Science & Politics 54 (3): 473–78. https://doi.org/10.1017/S1049096521000287.

———. 2021c. "Introduction: Case Studies in Transparent Qualitative Research." Qualitative & Multi-Method Research 19 (1): 6–9. https://doi.org/10.5281/zenodo.5495548.

Karcher, Sebastian, Dessislava Kirilova, Christiane Pagé, and Nic Weber. 2021. "How Data Curation Enables Epistemically Responsible Reuse of Qualitative Data." The Qualitative Report 26 (6): 1996–2010. https://doi.org/10.46743/2160-3715/2021.5012.

Weber, Nicholas, and Sebastian Karcher. 2020. "Seeking Justification: How Expert Reviewers Validate Empirical Claims with Data Annotations." In Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020, 227–34. JCDL '20. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3383583.3398537.

Weber, Nicholas, Sebastian Karcher, and James Myers. 2020. "Open Source Tools for Scaling Data Curation at QDR." The Code4Lib Journal, no. 49 (August). https://journal.code4lib.org/articles/15436.

Alexander, Steven M., Kristal Jones, Nathan J. Bennett, Amber Budden, Michael Cox, Mercè Crosas, Edward T. Game,... Sebastian Karcher, ..., Nic Weber et al. 2019. "Qualitative Data Sharing and Synthesis for Sustainability Science." Nature Sustainability, November, 1–8. https://doi.org/10.1038/s41893-019-0434-8.

Karcher, Sebastian, and Sophia Lafferty-Hess. 2019. "An Epic Journey in Sharing: The Story of a Young Researcher's Journey to Share Her Data and the Information Professionals Who Tried to Help." IASSIST Quarterly 43 (1): 1–10. https://doi.org/10.29173/iq942.

Karcher, Sebastian, and Nicholas Weber. 2019. "Annotation for Transparent Inquiry: Transparent Data and Analysis for Qualitative Research." IASSIST Quarterly 43 (2): 1–9. https://doi.org/10.29173/iq959.

Fenner, Martin, Mercè Crosas, Jeffrey S. Grethe, David Kennedy, Henning Hermjakob, Phillippe Rocca-Serra, Gustavo Durand, ..., Sebastian Karcher, ... et al. 2019. "A Data Citation Roadmap for Scholarly Data Repositories." Scientific Data 6 (1): 28. https://doi.org/10.1038/s41597-019-0031-8.

Mannheimer, Sara, Amy Pienta, Dessislava Kirilova, Colin Elman, and Amber Wutich. 2018. "Qualitative Data Sharing: Data Repositories and Academic Libraries as Key Partners in Addressing Challenges." American Behavioral Scientist, June, 0002764218784991. https://doi.org/10.1177/0002764218784991.

Elman, Colin, Diana Kapiszewski, and Arthur Lupia. 2018. "Transparent Social Inquiry: Implications for Political Science." Annual Review of Political Science 21 (1): 29–47. https://doi.org/10.1146/annurev-polisci-091515-025429.

Kirilova, Dessi, and Sebastian Karcher. 2017. "Rethinking Data Sharing and Human Participant Protection in Social Science Research: Applications from the Qualitative Realm." Data Science Journal 16 (September). https://doi.org/10.5334/dsj-2017-043.

Karcher, Sebastian, Dessislava Kirilova, and Nicholas Weber. 2016. "Beyond the Matrix: Repository Services for Qualitative Data." IFLA Journal 42 (4): 292–302. https://doi.org/10.1177/0340035216672870.

Elman, Colin, and Colleen Dougherty Burton. 2016. "Research Cycles: Adding More Substance to the Spin." Perspectives on Politics 14 (4): 1067–70. https://doi.org/10.1017/S1537592716002991.

Elman, Colin, Diana Kapiszewski, and Dessislava Kirilova. 2015. "Learning through Research: Using Data to Train Undergraduates in Qualitative Methods." PS: Political Science & Politics 48 (1): 39–43. https://doi.org/10.1017/S1049096514001577.

Elman, Colin, and Diana Kapiszewski. 2014. "Data Access and Research Transparency in the Qualitative Tradition." PS: Political Science & Politics 47 (01): 43–47. https://doi.org/10.1017/S1049096513001777.

Kapiszewski, Diana, and Dessislava Kirilova. 2014. "Transparency in Qualitative Security Studies Research: Standards, Benefits, and Challenges." Security Studies 23 (4): 699–707. https://doi.org/10.1080/09636412.2014.970408.

Lupia, Arthur, and Colin Elman. 2014. "Openness in Political Science: Data Access and Research Transparency." PS: Political Science & Politics 47 (01): 19–42. https://doi.org/10.1017/S1049096513001716.

Elman, Colin, Diana Kapiszewski, and Lorena Vinuela. 2010. "Qualitative Data Archiving: Rewards and Challenges." PS: Political Science & Politics 43 (01): 23–27. https://doi.org/10.1017/S104909651099077X