

A GUIDE TO ACTIVE CITATION

Version 1.1 for Pilot Projects

(January 2013)

I. Overview

This memorandum offers concrete guidance on how to make the presentation and use of qualitative data in published scholarship more transparent. This is an important goal, whether the data are employed for description, narrative, interpretation, analysis, or causal inference.¹ The document aims to provide a strategy to accommodate evolving disciplinary standards for data access and research transparency in ways that honor the qualitative research tradition and that guard against excessive or inappropriate homogenization of social science practices.²

This memorandum focuses on one approach to improving data access and research transparency in published qualitative work: *active citation*.³ This efficient and effective method envisages that claims (and data sources) that are central to an author's argument, and claims (and data sources) that are controversial and contestable within the literature the author is addressing, will be backed by "rigorous, annotated citations hyperlinked to the sources themselves." Active citation is thus a technologically enabled version of existing footnoting, citation, and appendix practice. It does not require qualitative political scientists to conduct research fundamentally differently than they do today, but rather simply encourages them to be clearer and more systematic in explaining what they do and why they do it.

This version of *A Guide to Active Citation* is provisional, and specifically tailored for use by scholars preparing "pilot projects" for the Qualitative Data Repository (QDR) at Syracuse University. While we have tried to work out as many of the details as we could, gaps remain. We hope that the authors of the pilot projects will help us to further develop and improve the process of active citation, as well as to sharpen the description and guidance for future users. Hence, as you proceed with activation, it will be very helpful if you could let us know where the instructions are vague, confusing, or lacking necessary information. We hope developing these pilot projects will be an intellectually engaging task, and we are very grateful for your assistance and patience.

II. Background: Emerging Standards for Research Transparency

Norms in the discipline of political science are evolving to encourage greater access to data and greater transparency in research practices. The American Political Science Association (APSA) is at the forefront of these changes. In October 2012, the APSA Council approved a formal

¹ A sister document, a "Qualitative Data Guide," supplements this document. It offers a brief set of guidelines for scholars who will be depositing data in the Qualitative Data Repository, outlining what constitutes a qualitative data collection and describing how to prepare qualitative data for sharing.

² On the issue of homogenization, see King, Keohane, and Verba (1994) as well as Goertz and Mahoney (2012).

³ Active citation was first proposed by Moravcsik (2010); see also Moravcsik (2012a).

amendment to the Association's *Guide to Professional Ethics in Political Science* (proposed by the Committee on Professional Ethics, Rights, and Freedoms) outlining revised principles for how scholars should present their research and the evidence upon which it is based.⁴

The revised *Guide* specifies that researchers “have an ethical obligation to facilitate the evaluation of their evidence-based knowledge claims through data access, production transparency, and analytic transparency so that their work can be tested or replicated.” These three key terms are defined as follows:

Data Access: Researchers making evidence-based knowledge claims should reference the data they used to make those claims. If these are data they themselves generated or collected, researchers should provide access to those data or explain why they cannot.

Production Transparency: Researchers providing access to data they themselves generated or collected should offer a full account of the procedures used to collect or generate the data.

Analytic Transparency: Researchers making evidence-based knowledge claims should provide a full account of how they draw their analytic conclusions from the data, i.e., clearly explicate the links connecting data to conclusions.

While increasing data access, production transparency and analytic transparency are now incumbent on all political scientists, there is also agreement that these general principles should be conceived and realized differently in the quantitative and qualitative research traditions.

III. How does Active Citation Further Data Access and Research Transparency?

The qualitative research tradition will be well served by augmenting existing citation practices with a more deliberate approach to the presentation and discussion of data. Active citation makes research more transparent and facilitates its substantiation—by making it easier for authors to show, and readers to see, the evidence on which authors rely, and how they rely on it. Hence, the method increases political scientists' ability to demonstrate the rigor and validity of their work, as well as to enhance its descriptive richness and interpretive depth. The aim of the practices described here is to make what is already good better, not to call into question the rigor or value of existing research, or to change the way scholars write books and articles. It seeks to encourage the broader use of scholarly “best practices” across the discipline.⁵

⁴ APSA (2008)

⁵ In doing so, active citation borrows existing “best practices” in scholarly fields such as legal academia, classics, policy analysis, historiography, and natural sciences, where rigorous attention to qualitative data and data analysis have long been a hallmark. Most electronic law reviews employ annotated, hyperlinked active citations to legal materials. Citations to original texts in classics journals are customarily hyperlinked through to sources. Citations to published research in natural science and medical journals are handled analogously. Some history journals have moved toward multi-media active citation, not just for text, but for scans of documents, visual and recorded material, and other items. As a result of controversy about the sourcing of their analyses prior to the Iraq War and other high-profile decisions, US intelligence agencies have also moved to active citation: even high-level analytic documents

The active citation standard builds upon traditional referencing techniques, which support evidence-based claims by providing information about sources in either footnotes or endnotes. Active citation leverages this traditional architecture by providing supplemental information in a “Transparency Appendix” (TRAP). Traditional practices already anticipate that evidence-based claims and related data sources will be referenced. Active citation suggests that where those evidence-based claims and data sources are central to an author’s overall argument, or where those evidence-based claims and data sources are controversial and contestable within the research community to which he or she is speaking, the author will provide supplemental information. This will be collected in the TRAP, which will be attached to articles and chapters, and which will store excerpts from the relevant sources, as well as annotations to those sources.

Active citation promises to improve research by enhancing data access, production transparency and analytic transparency, as we describe below in more detail. Active citation provides an efficient and immediate form of electronic data access for qualitative sources, without disturbing existing formats or page limits of scholarly work. Each source can be annotated as much as is appropriate, and additional discussion can be offered concerning analytic techniques, aggregation of findings, or empirical assessment, directly enhancing analytic transparency. Scholars may also strengthen production and analytic transparency by employing the TRAP to hyperlink appropriate sections of the main text to additional information about other aspects of the research project, including its overall trajectory, its research design and the methods used, and the data collection and selection procedures which were employed. In all these ways, active citation constitutes a very promising means of realizing disciplinary standards of data access and research transparency, while retaining and enhancing the richness and nuance of qualitative scholarship.

To be sure, not every reader will plumb the depths of every active citation. Many (perhaps even most) readers will be content with the main text of a paper, article or book, and perhaps a few of the footnotes. Activated research will be of most direct benefit to readers who are interested in investigating whether and how the information contained in cited sources supports authors’ evidence-based claims, and, in particular, to those who are interested in engaging, critiquing, and extending the particular line of research. Yet no matter how many scholars click open active footnotes, the emerging norm of active citation will have the collateral benefit of encouraging scholars to be more rigorous when making and supporting such claims – thus improving the overall quality of political science research.

A. Active Citation and Data Access

Active citation offers an effective and efficient means for qualitative social scientists to meet basic data access standards by linking specific arguments in the main text to data sources via

are now footnoted electronically all the way down to individual sources. Natural science journals make near universal use of appendices, often dozens or hundreds of pages long, to supply detailed data and background information on methods.

footnotes. Three characteristics of active citation account for its effectiveness and efficiency in doing so.

First, *by requiring precise references, active citation allows scholars to identify and locate sources*. The best traditional footnotes carefully identify supporting sources, and where the relevant evidence is located in those sources. However, all too often the information about particular sources and/or the location of evidence within them is incomplete or vague, making it difficult for other scholars to determine whether the cited evidence supports an author's claims. Even such basic elements as page numbers are often missing, as well as what aspect of the cited material supports the claim in the text. Some "scientific" citation methods do not permit scholars to display sufficient information to locate various types of primary and archival documents. Active citations address this problem by assuring sufficient information for readers to locate cited sources and to find the evidence to which an author is pointing to substantiate his or her claims, and by providing an excerpt of that evidence.

Second, *active citation places data directly before the reader*. Even where traditional footnotes precisely identify the location of a source and of the relevant evidence it contains, readers must fetch the materials for themselves. There is no tradition of (and no supporting infrastructure for) providing access to cited sources. Thus while precise citation makes it possible, in the abstract, for readers to access cited materials and evaluate to what degree they support an author's claims, in practice the transaction costs are often so high that the checking of empirical claims is a rare event. This is particularly important in cases where qualitative scholars generate or collect their own data, for example via interviews, archival research, or participant observation, rather than relying on existing datasets.⁶ Active citation addresses this problem by hyperlinking to sources, thereby dramatically reduce the costs of evaluating the extent to which an author's evidence substantiates his or her claims. With active citation, the evidence required to assess the *prima facie* empirical plausibility of any claim and to place it in a broader interpretive context is just one click away.

Third, *the active citation format facilitates re-use of data for replication or secondary analysis*. Scholars reading published or unpublished work often wish to reanalyze data to replicate or critique the analysis, or reuse data in order to extend the analysis or launch new research projects (secondary analysis). Reanalysis and reuse of this kind are commonplace in quantitative political science, in part because datasets are often readily accessible in electronic form. They remain rarer in qualitative traditions, in large part because the data are more costly and difficult to access. Active citation encourages reanalysis and reuse, because each source is not only available by direct hyperlink to a specific citation, but simultaneously resides with all other sources in the collective TRAP. Potential secondary users of evidence need not cull each source separately, but can immediately examine, download or otherwise manipulate the full body of evidence cited in an article or chapter for whatever scholarly purpose they wish—analogously to the way they could download an electronically accessible data set.

⁶ This is true even with published secondary sources, which may be difficult to find or be written in foreign languages. Primary source material may have been collected from multiple remote archives, for which a cumbersome process is required to attain the requisite permissions and similar linguistic barriers arise.

B. Active Citation and Production Transparency

The primary purpose of active citation is to clarify connections between individual footnoted claims and their supporting data sources. In order for the research design, descriptive and causal inferences, and interpretive choices to be clear to other scholars, however, they must understand how and why those particular data were generated or collected and why those data (and not others) were used in the analysis. The standard of *production transparency* requires authors to explain the origin, generation or collection, and selection of data supporting major empirical claims.

This is important in qualitative analysis, where the potential for particular biases (such as sampling bias) often looms large. One common form of bias results from “cherry-picking” evidence (i.e., over-representing evidence confirming a favored hypothesis). This can arise purposefully or inadvertently; and it can arise whether a scholar employs primary or secondary, oral or written, textual or visual sources.⁷ Such biases are generally difficult to spot because, in current practice, the absence of reference to evidence favoring other explanations is an error of omission invisible to all but the most expert scholar who already knows the data. While one can never eliminate entirely the risk of selection bias in data production—and in some types of work, it may not be desirable to seek to do so entirely—rendering data-gathering procedures more transparent should be a goal. Doing so encourages authors to be more self-conscious both about causal claims and about interpretations of subjective experiences, renders research results more legitimate and tractable, and facilitates scholarly debate and criticism. While some issues of production transparency are generally described in the text of the book or article, word limits and other space limitations, as well as the lack of a general expectation that they be dealt with, means that they are infrequently fully addressed.

Active citation provides a useful format in which to increase production transparency. A footnote to the discussion of data sampling and production in the research design section of the book or article can be hyperlinked to an annotation in the Transparency Index that supplements (rather than replaces or repeats) what is said in the text, providing additional context and background to the author’s research efforts. The author can describe in more detail the project’s empirical base, offer a holistic, synthetic discussion of the context of data collection, and elaborate the procedures he or she used to access, select, collect, generate, and/or capture data.

C. Active Citation and Analytic Transparency

Active citation’s greatest strength is its ability to facilitate *analytic transparency*. It does so by rendering patent the critical links between evidence-based description, narrative, interpretation, or causal inference in the text, and the particular data on which it is based. Crucially, by building upon traditional referencing practices, this transparency preserves the unique epistemic contributions made by qualitative research.

⁷ Lustick (1996).

Active citations link empirical claims in the text to the data sources authors use to support them. This improves analytic transparency in three ways.

First, *active citation links individual pieces of evidence to aspects of an inferential or interpretive analytical process*. Perhaps the most ubiquitous among the various approaches is process tracing, where scholars identify “causal process observations,” defined as “an insight or a piece of data that provides information about context or mechanism and contributes...leverage in causal inference.”⁸ Such data are not “variable scores to be assembled into a rectangular dataset,” but differing types of evidence employed to facilitate process-tracing or test specific counterfactual hypotheses about mechanisms within cases. Most often they comprise a series of observations linked to specific stages, aspects or elements of a narrative, process, or sequence.⁹ By placing such evidence in footnotes to the points in the text where relevant aspects, stages or elements of the causal process are discussed, the active citation format makes particularly transparent the way inferences are being drawn from data. It does so in a way that respects the particular characteristics of qualitative data.

Second, *active citation places arguments in a broader evidentiary context*. Rather than linking arguments to a citation (or, at best, a very short quotation), the active citation standard links arguments (through footnotes) to an excerpt from the cited source sufficient for a reader to judge whether the evidence it contains actually supports the description, narrative, interpretation, or causal inference in the main text. This offers some protection against taking evidence out of context and can offer the reader access to the specific cultural, strategic or social context from which particular social facts emerge. Social actors in different settings being studied “speak in their own voices” to readers.

Third, *active footnotes are annotated, offering researchers an opportunity to “show their work.”* Activated footnotes are generally accompanied by an annotation describing more precisely how the author used the data cited to arrive at the descriptive claim, interpretation, or causal inference discussed in the main text. They thus differ from traditional un-annotated citations (and, in particular, from increasingly widespread “scientific” citation styles), which simply juxtapose claims and references without any option to further explain the precise connection between them. In order to evaluate that connection in a traditional reference, other scholars must engage in “reverse engineering” – seeking to identify the link the author is implicitly positing. Yet, as Moravcsik (2012, 34) notes, in order to effectively evaluate evidence-based claims, readers need to know “exactly how and why the citation supports the textual point.” Further, active footnotes permit scholars to document a wider and richer range of descriptive, interpretive and causal arguments. Authors may use annotations to underscore the strength of an interpretive, descriptive or causal claim – or to acknowledge sources of complexity, uncertainty or ambiguity in such claims. As mentioned earlier, the active footnote format and the TRAP may also be employed to clarify methodological issues not fully explained in the text or footnotes.

⁸ Collier, Brady and Seawright (2010), 184. See also Mahoney (2012), Waldner (2012).

⁹ Goertz and Mahoney, p. 88-90. By contrast, in a quantitative dataset, data consist of a body of values of variables relating to aspects of outcomes that vary across cases. In a qualitative dataset, there is usually less attention to variation across cases but much more to establishing multiple observations of different links in the causal chain. The importance of each observation is contextual, depending on where it is situated.

IV. Standards and Logistics of Active Citation

This section examines practical issues and standards for research using active citation. It discusses standards for deciding the documents to which active citation applies; the citations within those documents that should be activated; the proper format for the TRAP; the format and standards for creating a reference citation, annotation and source excerpt; and the recommended way to add more generic methodological entries.

A. To what Research do Active Citation and Research Transparency Standards Apply?

Active citation can be used to facilitate data access, production transparency, and analytic transparency in all qualitative political science research. It will be particularly useful in meeting the new APSA standards for data access and research transparency in published scholarship.

With regard to timing, data supporting evidence-based claims in published scholarship must be made fully accessible – that is, active citation must be fully implemented – within one year of publication (although funders, journals, conferences, hiring committees, and publishers, may require earlier revelation of data). However, earlier revelation of qualitative data in the form of active citation may be preferable for logistical, intellectual and policy reasons. From a logistical standpoint, it may be costly for journal and book publishers and editors to publish work and then to revisit it a year later to add hyperlinks and sources. Similarly for authors, it is easier to implement active citation if information has been inputted from the start. From an intellectual standpoint, it is only possible to judge the *prima facie* rigor and plausibility of evidence-based claims if the data sources that underlie them are available. If they are not, journal and book reviewers and editors must positively assess and accept work for publication without being able to assess its true quality. This not only foregoes opportunity for critique and improvement, but risks embarrassment if the sources are revealed and found to be inadequate. From a policy perspective, qualitative researchers may assume less professional risk when sharing their data in this way, since only material authors actually cite must be revealed and in excerpted form.

B. When and How is Active Citation Employed?

We suggest four criteria that can be used to determine which footnotes should be activated in a published chapter or article, and consider different types of citations that might appear in published scholarly work (see Table 1).¹⁰

¹⁰ APSA's amended standards for data access and research transparency were developed for application to published research. Accordingly, the discussion in this section focuses on published books and articles. Although it is not a formal requirement, to the extent authors make evidence-based arguments in working papers and other unpublished material, we believe it is best practice to also use active citation in those formats.

One of the most difficult questions for deciding the new standards is determining which citations need to be activated. In practice the answer will be more than none, and fewer than all, requiring authors to select those references that require activation. We propose the following criteria for fully active citation:

- [a] **Centrality of the Evidence-based Claim:** Footnotes to evidence-based claims that are crucial building blocks to an author's overall argument or thesis should be activated;
- [b] **Importance of the Data Source:** Footnotes referencing crucial pieces of evidence underlying the author's overall argument or thesis should be activated;
- [c] **Contested or Controversial Nature of the Evidence-based Claim:** Footnotes to evidence-based claims that are contested or controversial within the research community, debate, or literature to which an author is speaking or on which he or she draws, or might be questioned by members of those communities, should be activated;¹¹
- [d] **Contested or Controversial Nature of the Data:** Footnotes referencing data sources that are controversial within the research community, debate, or literature to which the author is speaking and on which he or she draws, or whose validity or evidentiary value might be questioned by members of those communities, should be activated.

We recognize that these criteria have some associated difficulties. For example, all involve judgments by the author. We expect that these will be made on good faith, and also hope the criteria provide a framework in which the sourcing and annotation of a work can be challenged by other members of the research community. Similarly, we recognize that the degree to which a claim or data are controversial is subjective, and may be specific to a particular time (what is accepted at once moment may be called into question later). In this regard, however, these criteria do require that an author provide sourcing for claims that might be controversial not just within the discipline *to* which the article is directed (e.g. political science) but within any discipline *from* which it draws material. This offers some protection against cherry-picking results from other disciplines.¹² Overall, these criteria make it more likely that the activated citations will deliver useful transparency, while being mindful of the opportunity costs involved in active citation, and hence the pragmatic limitations on what it is reasonable to expect authors to undertake.

The fully active citation (with annotation and source) is recommended as the standard default format for any claim or data meeting one or more of the centrality/controversy criteria mentioned

¹¹ Consider a few examples. While the dates of President Obama's inaugurations may seem obvious, he in fact was sworn in on two different dates for each term, in part as a result of the complex constitutional consequences of Justice Roberts' having misspoken at his initial inauguration. The true population of China and other countries is the subject of considerable debate. The word count of the Declaration of Independence depends on which words one counts: signatures or not, headings or not. These ambiguities may or may not matter for any given study. A study of the constitutional history and politics of inaugurations might require source material on these points, whereas a study of inaugural speeches, for which the date is simply a peripheral fact, may not require an annotation or a source, or even a citation.

¹² Lustick (1996).

above. However, since not all citations will be fully activated, it is helpful to distinguish between four types of citations which might appear in scholarly work: (1) a traditional citation, (2) a fully active citation (with both annotation and source material); (3) an annotated traditional citation; and (4) an un-annotated active citation.

Under the active citation standard, we recommend and anticipate that traditional citations and fully active citations will be the most prevalent, perhaps augmented by very sparing use of annotated traditional citations.

Table 1: Four Types of Citations

		ANNOTATION PROVIDED	
		YES	NO
SOURCE PROVIDED	YES	Fully Active Citation	(Un-annotated Active Citation)
	NO	Annotated Traditional Citation	Traditional Citation

1. The ***traditional citation*** simply contains a reference citation, with neither annotation nor source material appended. Within the active citation standard, traditional citations remain appropriate when none of the criteria outlined above is met, and the footnote simply, for instance, cites alternative social science theories, provides methodological guidance, acknowledges previous work on the topic, offers general empirical background, discusses background examples the author does not intend to develop in detail, establishes the general importance or policy relevance of a study, to set the mood by providing epigrams, etc. By contrast, if the argument of the study in question specifically and centrally problematizes the veracity of these items, or rests on specific, questionable assumptions about them, an active citation is likely to be appropriate.
2. The ***fully active citation*** (with annotation and source) is recommended as the standard default format for any claim or data meeting one or more of the centrality/controversy criteria mentioned above. The footnote (in the format of the publication) is hyper-linked to a “source entry” in the TRAP containing a full citation, an annotation, and (presumptively, in the case of textual sources) an excerpt from the source at least 50-100 words long. This requirement holds whether the text is a primary document, a redaction or summary of a document, a transcript of an interview or focus group, an oral history, a set of ethnographic or participant-observer field notes, a diary and other personal record, a press clipping, a secondary publication, etc.. The source length requirement may be adjusted to meet human subject, intellectual property, and logistical constraints, as discussed in more detail below. It will often be of great benefit for the fully active citation to be accompanied by the underlying textual or non-textual source (e.g., a scanned document, audio or video recording, photograph, map, poster, artwork and other visual artifact, etc.).

3. The ***annotated traditional citation*** is simply a footnote (formatted as customary in the publication) containing a reference citation and annotation. This type of footnote, which has long offered scholars a means of elaborating the relationship between evidence and argument, has been used less extensively in political science in recent years as “scientific” citation and endnotes have supplanted it and word limits have grown tighter. When such footnotes are not permitted by a publication or space limitations restrict their use, the TRAP might be used for this purpose. An annotation without a source may be appropriate when the cited source documents an unambiguous and uncontested historical fact, for instance, but the relationship between the claim in the main text and the citation requires elaboration. In most cases where a causal claim is contestable, however, residual ambiguity will likely exist about the content, veracity or representativeness of the cited source. In such cases, it is preferable that the source be presented in a fully active citation.
4. The ***un-annotated active citation*** provides a reference citation (formatted as customary in the publication) and is hyperlinked to the TRAP, where the citation is repeated and an excerpt from the source is transcribed, without annotation. We believe this type of citation should be avoided in practice for several reasons. An appropriate case might seem to arise when a controversial or contestable evidence-based claim is advanced, to which the source material is critical, but interpretation of the source and its relationship to the claim are unambiguous and uncontested. Yet few claims requiring a source are without any ambiguity. Moreover, the absence of any explanation would mean that a subsequent scholar consulting the TRAP would remain ignorant of why the source was cited, without referring back to the article. Even where the meaning seems quite clear, a very short (e.g., one sentence) annotation (i.e., a fully active citation) is preferable.

C. Creating Source Entries in the TRAP

Engaging in active citation involves creating a TRAP – a standardized means to present information about data, how they were selected, and how interpretations and inferences were drawn from them. It assists scholars in efficiently conforming to standards of data access, production transparency and analytic transparency. The TRAP is normally written in tandem with a book chapter or article (although for pilot projects based on existing research it is obviously being generated retroactively). For the purposes of the pilot projects, the TRAP will reside in the QDR.

Before continuing, we wish to note that we understand that activating citations to legacy data (i.e., material collected for and employed in projects that have been completed) is difficult for several reasons. Authors may not have formally recorded their data-selection and data-collection procedures, or may not have retained their notes. Moreover, sampling techniques and choices, analytical and interpretive decisions, and trajectory of research are often driven by a complex mix of intellectual and pragmatic considerations, which may be difficult for scholars to reconstruct years later. We hope authors working with legacy data will do their best to recall the research process.

The TRAP includes “source” entries (treated in this sub-section) and “methodological” entries (treated in the next).

Here consider first source entries. Each citation being activated has one corresponding source entry, to which it is hyperlinked. Each source entry contains, in the following order: (1) an identification label; (2) a reference citation; (3) location information; (4) information on availability and shareability; (5) source material; (6) an annotation; (7) hyper-links from the citation in the article/chapter text to the “entry” and back again.

1. The Identification Label. Each source entry (and each methodological entry) in the TRAP contains a number locating the spot in the article or book chapter where the citation is located.¹³

- *For books with numbered footnotes or endnotes:* The identification label should include the chapter number, footnote number, and a letter for the source. For example, if several citations appear in Chapter 7, Footnote 24, the first item would be labeled 7-24-a, second item 7-24-b, and so on.
 - *For published articles:* The identification label should include the footnote number, and a letter for the source. For example, in Footnote 24, the first item would be 24-a, second item would be 24-b, and so on.
 - *For books and published articles with non-numbered citations (in the text or in endnotes referring to pages):* For books, the identification label should include the chapter, page number, and a letter for the source. For example, in Chapter 7, page 300, the first item would be 7-300-a, the second item would be 7-300-b, etc. For other works, the first item cited on p. 212 would be 212-a, second item would be 212-b, and so on. The page number refers to the point in the main text where the citation appears or the text to which the citation refers, not the location of an endnote. (For those who are citing electronic books without page numbers, we are still developing a proper format.)
- 2. The Reference Citation.** Each source entry begins with a precise and complete reference citation, sufficient to permit readers to locate the cited source and find within it the precise passage of textual evidence to which an author is referring (i.e., page numbers).
- 3. Location.** Sometimes additional information will be necessary to allow scholars to locate the source and find the relevant information within it. For instance, for primary archival sources, information about the archive, collection, and number of the box in which the document was found should be noted. This information should be included even where it replicates what is already in the main text or relevant footnote of the article or chapter. This permits scholars to employ source material in the TRAP without referring back to the article.

¹³ Once the numbering has been completed for a large number of citations, it may be difficult and time-consuming to revise. These labels need only be included in the published versions of articles and books. Eventually, we expect that dedicated software will be developed to ease this task.

4. **Availability and Shareability.** Scholars should indicate whether they are in possession of a hard copy of the source and if they are in a position to post it to the QDR in tandem with the TRAP. If they can, they should also confirm that it can be shared per the rules under which it was collected and relevant copyright law. In this section of the source entry, scholars can also provide a hyperlink to the full source (if it is available online).
5. **The Source.** For the purposes of the pilot projects, textual sources excerpted in an active footnote must in all cases be transcribed and entered in the source entry as text. This makes the TRAP a self-sufficient document containing both excerpts from and annotations to all sources; renders the source material within it easily readable and fully searchable for scholars interested in locating, examining, reusing or downloading source material; and eliminates problems of outdated, inoperative, unstable external links, as well as unreadable external documents, which bedevil efforts to hyperlink footnotes in other fields.¹⁴

How long should the source excerpt be? For all sources, textual and non-textual, four general standards govern the proper amount of the source to be entered:

- (i) Enough of the source should be entered to clarify the nature of the evidence and provide sufficient context to interpret that evidence properly and with nuance.
- (ii) The excerpt should not violate intellectual property or human subject protections.
- (iii) The length of the excerpt should not impose a logistically prohibitive burden.
- (iv) Within these constraints, scholars should provide as much evidence as possible.

We recommend a *presumptive minimum length of 50-100 words/characters for textual sources*. This length is often minimally sufficient to convey the general meaning and context, lies within current scholarly practices with regard to “fair use” of most types of intellectual property, and imposes a manageable burden with regard to transcription and sanitizing sources for human subject purposes. Scholars enjoy discretion in interpreting this presumptive norm, since circumstances will surely arise where it is appropriate to fall short or not cite anything at all.¹⁵ When this occurs, the author must explain in the annotation to the entry why the excerpt diverges from the presumptive standard. Exceeding the standard is encouraged, and requires no explanation—assuming that human subject and intellectual property requirements are met, and the material is relevant.

¹⁴ At the present time, including source material as text directly in the TRAP is the best way to assure an adequate level of data access and analytic transparency, and to create a comprehensive and fully searchable location for scholars interested in examining or reusing source material. With technology and broader data practices moving at a rapid pace, however, it is possible that this will change in the future for some sources. We anticipate the need for several characteristics before a reliance on external links will be possible, including: permanent and stable location; swift means to examine, search and download single and multiple annotations and sources; uniformity of treatment and format between hyperlinked and traditional sources; translation and legibility; and a means for authors to highlight or excerpt critical passages or sections in a larger document or website. Until such a software solution exists, excerpting text in the TRAP remains the best way to implement active citation. Having provided that text in the TRAP, however, scholars are strongly encouraged to supplement that text with a link to external on-line materials if they are available, even if there is some risk of instability.

¹⁵ For example, certain types of intellectual property, such as poems or art works, are often subject to more stringent “fair use” intellectual property restrictions.

Are there alternatives to a Verbatim Textual Citation? In some cases, an author may be unable to provide a full verbatim text of the presumptive length. For example, an author may not possess a source in verbatim form (e.g., only interview notes exist, or no notes were taken), it may be lost, or may not be shareable due to human subjects or copyright concerns. In such cases, the author should offer an explanation, and provide as much material as is appropriate and feasible, for example, research notes, or a substantial redaction or summary of the information on which the claim in the text is based.

What about sources in foreign languages? If the source is in a foreign language, it must be presented both in the foreign language and in translation in the language of the book or article. The name, date and reference information (if any) of the translation should be given. Access to the original language enhances research transparency. We recognize that this imposes a logistical limit on length, and thus the translation requirement applies only to the presumptive minimum source text; the author may apply it to additional text, scans, or other materials that are provided at his or her discretion.

What about non-textual sources? The general principles governing the conditions under which active citation should be employed and how much should be provided can also be met by non-textual sources, including visual representations, music, and such. Since this is a work in progress, there is not yet any presumptive minimum standard analogous to the 50-100 word rule. For a discussion of the mechanical steps for including these materials with pilot projects, see below on page 17.

What about permissions and intellectual property issues? We expect that for most sources, the minimum requirement will fall under “fair use” for non-profit purposes and will not require special permission. However, we are still formulating firm guidelines on this issue. For the purpose of pilot projects, the Qualitative Data Repository will try obtain any necessary permissions from authors, editors and publishers to reproduce the text, references, and sources. Where this is not possible, we will advise use of as much material as is permitted by fair use.

May one supply additional data? Once an author has met the basic minimum standard of source text, he or she may add additional text in the entry, or include a link to external on-line materials if they are available.¹⁶ He or she may also provide a scan of all or part of the original source – be it a primary document or secondary source; transcripts of interviews, focus groups, or oral histories; ethnographic or participant-observer field notes; diaries and other personal records; press clippings; research notes; photographs, maps, posters and other representational work, or artwork. The author may also provide copies or links to non-textual materials, such as audio or audiovisual recordings, artwork, maps, diagrams and other representation material, and other items. (For a discussion of the mechanical steps for including these materials with pilot projects, see below on page 17.)

¹⁶ As mentioned in footnote 15 above, at this time we recommend that any source needed to activate a citation be provided in the TRAP itself, with external links reserved for supplemental data.

6. **The Annotation.** Active citation requires that an annotation accompany most citations. The annotation specifies how the cited evidence relates to and supports the inference or interpretation drawn in the main text and/or footnote. Some annotations may simply amount to a “plain reading” of the source data, while substantiating other claims may require that the scholar explain how he is using certain information such that it amounts to evidence for the claim, or describe the complex synthesis of multiple sources.

What exactly is in the annotation? Annotations describe the analytic path between cited data and evidence-based claims in the text. The precise content of an annotation depends on the type of claim the author seeks to substantiate; the particular principles, logics, or rules of the analytic methods he or she is employing; the type of data; and the extent of any ambiguity or complexity in the evidence. Annotations might highlight both confirming and disconfirming evidence, and/or point out subtleties, nuances, contradictions and ambiguities in the evidence. If the claims in the text are contested, the annotation may discuss evidence that *refutes* potential criticisms, potentially referencing multiple and competing sources not cited in the text. In the context of legacy footnotes activated subsequent to publication, annotations might be used to add information that emerged after publication.

Because this is a project in development, it is not yet possible to offer an exhaustive list of annotation types; indeed, we hope to develop a more complete list with the help of the pilot projects. Nonetheless, in general terms, we can distinguish between annotations supporting descriptive claims, and annotations supporting causal claims.

Annotations supporting *descriptive claims* supplement the text by making clear how the author is interpreting the cited data such that they are evidence of a particular descriptive assertion. Such annotations can substantiate a range of descriptive assertions, from simple statements of fact (“a” occurred on a particular date) to more nuanced or contested descriptive or interpretive assertions concerning the place or role of a particular element or node in a narrative, or the relationship between two elements or nodes, for instance, “a” preceded “b;” “a” followed “b;” “a” happened in a certain way; “a” was unusual; “a” was important; “a” meant a certain thing; “a” was interpreted in a particular way by a community, etc.

Annotations supporting *causal claims* supplement the text by making clear how the evidence shows that the asserted explanatory relationship holds. The specific content of these annotations depends on the analytic methods an author is employing and their particular principles, logics, or rules for drawing causal inferences, and the type of data he is using. To give just three examples, these annotations may show how “a” was necessary for “b;” how “a” was sufficient for “b;” or that a causal mechanism was operating in the way the author asserts. Put differently, the author’s methods “set the bar” for what he must do in order to make a persuasive causal argument; the degree to which his presentation and discussion of evidence (in the text and annotation) clear that bar determines the level of confidence readers should have in his causal claims.

How long should annotations be? Annotations need not be long, but there is no hard and fast rule. The appropriate length and detail may range from a one-sentence statement of the main

point to a detailed acknowledgement of controversy, ambiguity, or complexity in the relationship between text and source. If plain reading of the text, footnote, and source material make self-evident how the source relates to the interpretation or inference in main text, an extremely short annotation pointing to a critical passage or underscoring a relatively simple interpretation may be sufficient.¹⁷

What about multiple sources supporting a single argument? While under normal circumstances each citation receives its own source entry, sometimes a number of cited sources are used to support one claim in the text. Sometimes these sources are found in a single footnote, sometimes in sequential ones. If all such sources have the same relationship to the textual argument, or if one annotation serves to elaborate how they collectively support the underlying argument, then the entire annotation need be entered only for the first source. Each source thereafter may simply be annotated: “See annotation under Source Entry X above.”

7. **The Hyperlinks.** Each source entry contains a hyperlink back to the reference in the text or footnote. For the pilot projects, the Qualitative Data Repository will manually add hyperlinks to those online versions of the publications. Our experience with the pilot projects will help us to develop a standardized approach to locating and formatting these hyperlinks.¹⁸

What happens to hyperlinks in a hard-copy version of active citation? The active citation is an electronic format, but it can be employed also in a hard-copy format. The only difference is that the hyperlinks will not function and one must move manually from the main text, citations, to the TRAP, and back again, by flipping pages. To facilitate this, authors who expect that their papers will be read in hard copy should consider making sure that their source entries contain full labeling information about the location of the citation in the work (see C2 above on labeling), and, if possible within the format of the publication, referencing the sequential number of the source in the footnote, endnote or citation references.¹⁹

D. Creating Methodological Entries in the TRAP

Scholarly books and articles in social science typically include a research design indicating and justifying, among other things, data collection practices and analytic methods employed. The analytic methods that are typically employed in a qualitative research design (or a multi-method design with a significant qualitative component) might include process tracing, counterfactual

¹⁷ The author may feel that no annotation is needed to achieve analytic transparency, because a source alone is adequate. Yet this is often misleading. Situations in which no possible ambiguity exists, even in the mind of a reader without background knowledge, are rare. The recommended default minimum is thus a very short annotation.

¹⁸ Andrew Moravcsik, the inventor of active citation, has provided a very helpful document explaining one approach to formatting active citations and the Transparency Index in MSWord. See Moravcsik (2012b).

¹⁹ As suggested by footnote 15 above, the use of hardcopy versions of TRAP will make most sense when the sources are included as text in the appendix. To the extent that emphasis shifts over time to hyperlinking to external sources, we anticipate that hardcopy versions will become correspondingly less useful. Of course this phenomenon is not limited to active citation, but extends to any comparison of online and hardcopy versions of material in, for example, the digital humanities.

analysis, pattern matching, congruence testing, small-n comparison, or *fs*/QCA and other set theory approaches. These different methods imply different approaches to (and thus different rules for) drawing descriptive or causal inferences, and each relies on a set of underlying principles, logics, or rules which justify its use and support the claim that properly employing the method facilitates drawing valid descriptive and causal inferences.

While methodologists may differ concerning *which* rules to follow, most agree that following a particular set provides scholars with more inferential leverage than they would otherwise have. The APSA standard likewise prescribes no epistemology or methodology; it simply requires that whatever approach is chosen, analysis be conducted in as transparent a manner as possible, consistent with the general rules that attend the interpretive or inferential approach the researcher is using.

Methodological Entries in the TRAP supplement information provided in the text and footnotes, offering scholars an opportunity to clarify their research design; provide additional information about the procedures used to select, analyze, or interpret evidence and data; and illustrate that they have followed the rules that attend the methodology they have employed. In contrast to source entries, methodological entries contain only an annotation, but no source material. We imagine at least three basic types of methodological entry.

The first supplements the discussion in the text and footnotes about *the formation of theories, hypotheses, interpretations and narratives*. How did the analyst select theories, and generate, evaluate, and modify hypotheses, interpretations or narratives? By what means are observable implications derived from theories or interpretations? While these issues are normally treated in the background, literature review, theory, and hypotheses formation sections of published work, to the degree that the author believes production or analytic transparency can be enhanced by supplementing what is offered there, he can do so via a methodological entry in the TRAP.

The second type of methodological entry supplements the discussion of *data collection* in the existing text and footnotes. In much of political science scholarship, readers only see the evidence that is finally cited, raising the possibility of purposeful or inadvertent “confirmation bias.” While one cannot eliminate this possibility, production transparency requires that authors describe their projects’ empirical base and offer a holistic, synthetic discussion of the context of data collection, and the research design they employed to access, select, collect, generate, and/or capture data. What is the overall field of potential evidence and what justifies decisions to select or sample particular evidence to examine, to analyze in detail, and to cite?²⁰ How was the case

²⁰ This section might explain, for example, how the author moved from the “data field” (all of the information that is potentially relevant to the question under investigation) to “consulted data” (data the author considered and consulted while conducting research, formulating the question and basic claims) to “heuristic data” (data the author considers part of the relevant sample formulating an description, narrative, interpretation, or test of particular claims) and finally to “cited data” (data actually referenced in published work). For example, if one were studying a British foreign policy decision, one might indicate that one consulted the private papers and memoirs of the Foreign Secretary, Hansard (the House of Commons official report), and interviewed several former and current government officials. One would also justify why the particular private papers, reports, and interview respondents are relevant to one’s question, why they can be seen as an unbiased sample of a larger body of evidence, consider the biases that might arise from not examining the views of other ministers, lower officials, parliamentarians, social actors, and

(or cases) selected, and what are the implications for internal and external validity, or for cross-cultural interpretation? What instruments and procedures were employed to collect data? If interactive data-collection techniques such as interviews were employed, what procedures were employed to identify and access interview respondents, what terms of confidentiality and informed consent were employed; what interview protocols, topic guides and interviewer instructions were issued? Were data photographed, scanned, photocopied, recorded, summarized in notes, or otherwise stored? What relevant data were not or could not be consulted and what is there potential effect on the analysis? How were data prepared for sharing? Authors can enhance research transparency by providing information on these issues beyond what appears in the text or footnotes of their scholarly work in a methodological entry in the TRAP.²¹

The third type of Methodological Entry enhances discussion of *data analysis* by supplementing discussion in the text and footnotes about how inferences are drawn and claims are made from data. Individual source entries in the TRAP provide analytic transparency for individual pieces of evidence. Yet this still leaves open the question of why and how these individual findings are aggregated into claims about the confirmation and disconfirmation of hypotheses, and how these conclusions are in turn aggregated into claims about the confirmation and disconfirmation of broader theories. This is particularly complex since in most qualitative models, the hypotheses in question may correspond to different stages or aspects of a causal chain, rather than simply being different variables contributing to a common effect. Discussion of these broad issues of analytic transparency, which may often take place in the text or footnotes, can be supplemented by additional remarks in a methodological entry in the TRAP.

In sum, methodological entries in the TRAP offer an additional option for enhancing production and analytic transparency, should authors wish to be more explicit about general methodological issues than space elsewhere permits. How detailed each section of the TRAP will be depends upon the type of research project and publication, and the researcher's ambition with regard to rigor and precision.²² There is no single rule. While greater transparency may facilitate engagement with scholarly work, social scientists should not set unrealistic standards, letting the best become the enemy of the good. Data-collection practices, and the broader research process, are often messy, and discussions thereof in the TRAP will have both practical and methodological aspects.

other actors, explain how one interpreted or coded the evidence from these sources, and account for the selection of particular evidence for citation.

²¹ In the natural sciences, one can retrospectively trace the evolution of the project through time. This is essentially the function of the experimental "lab notebook." Notebook in hand, a researcher need only distill and systematically synthesize its content in order to describe how he or she generated and evaluated interpretations, data and inferences, and arrived at conclusions. While describing this entire process from beginning to end is not yet realistic as a universal expectation for political scientists (and is not a requirement here), it should be considered an aspirational goal. The more carefully political scientists keep track, contemporaneously with their research (or at least as near as possible), of the choices they make, the procedures and practices they used, and the sequence in which they employed them, the better able they will be to achieve production and analytic transparency.

²² For a project replicating an existing study in order to question the data analysis, or applying a set of established and consensual theories, hypotheses and inferences to new data, the focus may be largely on data collection and data analysis. A study that adds an original theory or set of hypotheses to an existing debate about a case may focus more on theory and hypotheses formation.

IV Additional Information Regarding Pilot Projects

For each pilot project, we will establish a Dropbox folder and share it with the author. Ultimately, pilot project authors will upload their completed TRAP to the Dropbox folder.

For both pragmatic and sociological reasons, the general Active Citation standard prescribes somewhat minimal practices for achieving data access and research transparency. However, it may be advisable for the purposes of the pilot projects to set slightly higher standards in order to illustrate the significant potential of Active Citation for achieving all three DA-RT goals – increasing data access, production transparency, and analytic transparency. More data and information increase the payoff with regard to each goal of activating book chapters or articles.

As noted previously, pilot projects will be displayed via the QDR. The QDR is also available to host any supporting material (e.g., unpublished primary sources, published primary sources, primary sources cited in secondary sources, secondary sources, or other research materials)²³ to which authors have access (or to which we might be able to get access), and/or any research instruments that were used in the analysis.

If you are in a position to send the document, please also confirm that it can be shared per human subjects agreements, the rules of the source from which you collected it, and relevant copyright law.

Underlying sources may be provided in various formats: textual, digital (e.g., word-processed, databases, spreadsheets), audio (digital and non-digital), graphical, video, and photographic (digital and non-digital). To the extent possible, QDR will hyperlink the activated citations to the underlying source.

If the source is in hard-copy form (e.g., newspaper clippings, typed or hand-written notes, photographs, documents) please provide a pdf of the item. (Please do so even if you cannot publicly share the item, as we will use it to check the redaction, quotation, or summary you provide for accuracy.) If the source is a secondary source and you have access to it, please provide a scan of the pages included in the citation.

Please name the pdf document using the following convention:

YourLastName_sourcelabel_AuthorOfSource.

So, for example, were Professor Hammond sending us the third document cited in footnote 24 to chapter 7, and the source was an article by Theda Skocpol, the file name would read:

Hammond_7-24-c_Skocpol

²³ This Guide's sister document, a "Qualitative Data Guide," has additional information on types of qualitative data and how to prepare them for sharing.

If the document has no author, please instead indicate where the document was found. Thus, instead:

Hammond_7-24-c_ReaganPresLibrary

BIBLIOGRAPHY

- American Political Science Association. 2008. *Guide to Professional Ethics in Political Science*. (Changes proposed by the Committee on Professional Ethics, Rights, and Freedoms, July 2012 available at: <https://www.apsanet.org/content_2483.cfm>.)
- Goertz, Gary and James Mahoney. 2012. *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences*. Princeton: Princeton University Press.
- Lustick, Ian. 1996. "History, Historiography, and Political Science: Multiple Historical Records and the Problem of Selection Bias," *American Political Science Review* 90(3): 605-618.
- Mahoney, James. 2012. "The Logic of Process Tracing Tests in the Social Sciences." *Sociological Methods & Research* 41(4): 570-597
- Moravcsik, Andrew. 2010. "Active citation: A precondition for replicable qualitative research." *PS: Political Science and Politics* no. 43 (1):29-35.
- Moravcsik, Andrew. 2012a. "Active Citation and Qualitative Political Science." *Qualitative and Multi-method Research (Newsletter of the Qualitative and Multi-Method Research section of the American Political Science Association)*. Spring: 33-37.
- Moravcsik (2012b) at http://www.princeton.edu/~amoravcs/library/active_citations.docx.
- Qualitative Data Repository. ND. "Qualitative Data Guide."
- Waldner, David. 2012. "Process Tracing and Causal Mechanisms." In Harold Kincaid, Ed., *The Oxford Handbook of Philosophy of Social Science*. New York: Oxford University Press, pp. 65-84.