

A GUIDE TO ANNOTATION FOR TRANSPARENT INQUIRY (ATI)¹
Version 1.0 (prepared for the ATI Pilot Working Group, Workshop #1)²
(December 2017)

Colin Elman (Syracuse University / QDR)
Diana Kapiszewski (Georgetown University / QDR)
Andrew Moravcsik (Princeton University)
Sebastian Karcher (Syracuse University / QDR)

I. Overview

This Guide introduces a new approach to openness in digitally published qualitative and multi-method research. *Annotation for Transparent Inquiry* (ATI) helps scholars to be more transparent about the generation and analysis of the data that underpin their publications, and to make that evidentiary base itself more accessible. These are important goals, whether the data are employed for causal or descriptive inference, or interpretation.

The Guide discusses the conceptual underpinnings of ATI, considers how ATI enhances transparency in qualitative and multi-method research in the health and social sciences, and offers suggestions with regard to employing ATI. Authors can also access a very brief [introduction to ATI](#), and [more information](#) about the technical aspects of using ATI, on the website of the Qualitative Data Repository ([QDR](#)).

II. Background: From Active Citation to Annotation for Transparent Inquiry

ATI builds on “active citation” (AC), an existing approach to transparency in qualitative research pioneered by Moravcsik (e.g., 2010, 2012a, 2012b, 2014a, 2014b, 2014c, 2016). AC rests on the premise that providing evidence, information, and interpretative/analytical commentary that go beyond what is included in the text of a publication (potentially even providing links to underlying data sources) can help render more transparent the *prima facie* basis of contestable knowledge-based claims.

AC recommends hyperlinking citations that back contestable empirical claims to entries in an appendix or in other similar supplementary materials. Each entry contains: (1) a brief textual excerpt or summary of the cited source, (2) an explanation of how the source is interpreted to support the textual point in question, (3) a full citation, if that does not appear in the text; and (4) optionally (only if ethically, legally and logistically feasible) a link to or copy of the source itself. In the first entry, authors can provide general information about how data were collected, selected and analyzed. In this way, AC offers a compelling framework in which authors can use

¹ We gratefully acknowledge helpful comments on the Guide to Active Citation (a precursor document to the present guide) from Louise Corti (Associate Director UK Data Service) and Dessislava Kirilova (Curation Specialist, Qualitative Data Repository).

² QDR is holding workshops on ATI in February and September 2018, both supported by generous funding from the Robert Wood Johnson Foundation. QDR is partnering on this initiative with Cambridge University Press, and with the software company Hypothes.is.

digital means to provide supplemental materials in a format that matches the traditional epistemic structure of inferences and interpretations in qualitative research.

A consensus in discussions of transparency exists that supplementary digital material should not be housed on personal websites but in more stable venues. A range of alternatives exist, including journal websites. AC remains agnostic as to precisely where and how supplemental digital materials should be stored and delivered.³ ATI, by contrast, holds that supporting data should be stored in venues, such as data repositories, that can meet modern data standards, including that data be FAIR (findable, accessible, interoperable, and reusable).⁴ FAIR data are more useful for transparency purposes, secondary analysis, and teaching. ATI seeks to meet these standards by using open annotation technology, which facilitates the generating, sharing, and discovery of digital annotations across the web. The annotations authors create when using ATI appear as an overlay to the web page where the research is published, but the annotations (including any underlying source material) are curated and preserved by QDR so as to render them FAIR.

AC and ATI do not require qualitative and multi-method researchers to change fundamentally how or why they conduct research. To the contrary, both provide ways for researchers to explain more clearly and systematically what they do and why they do it. They both allow scholars using text-based qualitative methods (and scholars within particular research communities within it) to highlight and enhance the distinctive elements of their particular research traditions, thereby avoiding the homogenization of social science practices.⁵ Moreover, they do so in a flexible way. For both AC and ATI, decisions about how, when, and where they should be employed in a manuscript remain up to authors, who possess the most detailed knowledge of the ethical and intellectual appropriateness and risks of doing so.

III. Why Annotation? Emerging Views on Transparency

Norms across the health and social sciences are evolving to encourage greater access to data and more transparency in research practices so that scholarly work can be more easily understood and evaluated. Different disciplines have adopted different terminology to describe these imperatives. We find three key terms adopted in political science to be particularly useful (see American Political Science Association 2012, 9-10), and employ them in this Guide:

“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.

³ Hyperlinked material might appear, for example, as an appendix or supplementary materials file on the journal publisher’s website.

⁴ As Kratz and Strasser (2014, 5) note, “data can be hosted by the journal as supplementary material or deposited in a third-party repository. The trend is away from supplemental materials because repositories are considered to be better suited to ensure long-term preservation and access to the data.”

⁵ On the issue of homogenization, see King, Keohane, and Verba (1994) as well as Goertz and Mahoney (2012).

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.”

In political science and elsewhere, scholars broadly agree that while these general principles are generally applicable, they should be realized differently in different research traditions and in varied circumstances. The rest of this Guide focuses on specific challenges and practices in qualitative research traditions.

IV. How ATI Advances Data Access, and Production & Analytic Transparency

Traditional practices of qualitative analysis in the health and social sciences and in the humanities use citations to link evidence-based claims to the data sources that support them. Recent technological advances offer the opportunity to augment existing citation practices to make that evidence more accessible, and to include more detailed discussions of how evidence was generated and how it supports arguments. That said, ATI is designed to be consistent with existing approaches to both doing and representing qualitative research. It is intended to help researchers demonstrate the strengths of their research, not to fundamentally change the way scholars write books and articles.

ATI empowers authors to demonstrate the richness, rigor and validity of their inferences and interpretations.⁶ Precisely how ATI can best be used to increase the transparency of qualitative research varies across research projects. Indeed, one goal of QDR’s ATI Initiative—the project for which this version of the Guide was written—is to develop a clear typology of the uses to which ATI can be put. To offer some examples, scholars can use annotation to elaborate on and support more rigorously evidence-based claims that are central to their overall argument, or controversial and contestable within the research community to which they are speaking. Scholars can also use annotation to provide more nuanced discussions of interpretations and inferences when word-count limitations complicate doing so in the main text. Annotations can also be used to describe the techniques used to generate or analyze evidence, and the aggregation of findings. Employed in any of these ways, annotations increase transparency.

Each annotation is linked to a particular passage in the manuscript being annotated, and each includes one or more of the following:

- Full citation to the underlying source(s) (when these are not in the publication’s bibliography); where relevant, supplementary information about the source’s location;
- Source excerpt(s): a quote (or redaction) from a textual source (including the transcription of handwritten or audiovisual material), typically 100 to 150 words;
- Source excerpt translation(s): a translation (and its source) if the excerpt is not in English;
- Analytic note: information contextualizing the source(s), and/or discussion of how the relevant source(s) were collected, how data were generated, how that source/data support conclusions or claims in the annotated passage;

⁶ For a detailed discussion of the essential elements of active citation and its advantages, consistent with this section on ATI, see Moravcsik (2014c).

- Data source: the file name(s) of the underlying digital data source(s) linked to the source(s) themselves when these are digital and can be shared ethically and legally.

An “ATI Data Supplement” for a particular manuscript comprises the set of digital annotations that the scholar created, as well as an “ATI Data Overview.” This overview, of approximately 1,000 words in length, discusses the data generation procedures that were employed, and how the analysis attends to the rules of inference or interpretation that underlie the qualitative methods employed. Because this overview supplements the manuscript, it need not repeat anything that is already included there. When publications have a bibliography or list of references, this is also included in the overview.

In short, ATI can make immediately available to readers – with just one click – the *prima facie* basis for a published empirical claim, and discussion of how it was generated and/or analyzed. Importantly, it does so without disturbing existing formats or page limits of scholarly work.

ATI benefits authors of qualitative and multi-method research in the health and social sciences in multiple ways. It encourages them to be more careful and precise when making and supporting evidence-based arguments, permits them to display critical evidence supporting their claims (thus augmenting the rigor and value of their research), and helps them to meet transparency standards while retaining and enhancing the richness and nuance of qualitative scholarship.

ATI also benefits readers of such scholarship. To be sure, not every reader will plumb the depths of every annotation. Many scholars will be content with reading the main text of a published work, and perhaps a few annotations. For them, the experience of reading a scholarly work will remain essentially unchanged. Annotation will be of great utility, however, to scholars seeking a richer account of an author’s descriptive and interpretive claims or wishing to investigate whether the information contained in cited sources supports evidence-based claims. A central benefit of ATI is reducing the transaction costs of engaging in such evaluation.

We now explain specifically how ATI furthers three core openness goals: data access, production transparency, and analytic transparency.

ATI and Data Access

Annotation helps qualitative researchers to meet basic data access standards by connecting specific conclusions or claims in the main text of a publication directly to data that support them via digital annotations. Doing so opens serious engagement in qualitative debates to a wider body of scholars, not just a few scholars who already “know the sources.” How does ATI do so?

First, by requiring precise citations, annotation encourages scholars to identify the sources that underlie their claims and clarify where the relevant evidence is located in those sources. For sources that may be difficult to find, for instance some types of primary and archival documents, ATI also allows scholars to describe in more detail the location of sources. Engaging in this type of clarification – all the while preserving existing citation formats – makes it easier for other scholars to find the evidence to which authors are pointing to substantiate their claims, and thus to determine whether cited evidence supports those claims.

Second, ATI places qualitative data directly before the reader. Traditional citation offers readers information *about*, or in the best cases *about the location of* key sources, e.g., primary documents (or redactions/summaries thereof); transcripts from interviews, focus groups, or oral histories; ethnographic or participant-observer field notes; diaries and other personal records; press clippings and so on. Yet sources often remain inaccessible to readers. Sometime sources are not publicly available, as in the case of transcripts and field notes. Other times the challenge is logistical: readers are unlikely to visit several archives to re-consult documents. This lack of availability makes evaluation of the extent to which those sources support an author's claims impossible. Put differently, the transaction costs of accessing data underlying published scholarship based on qualitative evidence are often so high that empirical claims are not checked.

ATI allows scholars to provide excerpts from the sources they use (and ideally the sources themselves), thereby dramatically reducing the costs of evaluating the extent to which an author's evidence substantiates their claims. Facilitating readers' access to underlying sources is particularly important when qualitative scholars employ primary sources or generate their own data (e.g., through interviews) rather than relying on existing datasets.⁷ With ATI, readers have before them the evidence required to assess the *prima facie* empirical plausibility of any claim, and often to place it in a broader interpretive context.⁸

ATI and Production Transparency

Production transparency requires authors to explain the origin and generation of data they cite in their publications. In order for an author's research design, her descriptive and causal inferences, and her interpretive choices to be clear, readers must be able to understand how and why those *particular* data (and not others) were generated or collected, how and why they were used in the analysis, and why they (and not others) were cited.

Some aspects of production transparency are generally addressed in the text of books and articles, and using ATI does not require that this information be repeated. However, given word limits and other constraints, scholars are seldom able to fully explain their data-generation procedures. Scholars using ATI have two means to clarify these points. First, scholars can provide in the "ATI Data Overview." Here authors can offer a holistic, synthetic discussion of the context of data collection, and elaborate on the procedures they used to access, collect, generate, capture, prepare, and select data for inclusion as evidence. Second, authors can use discrete annotations to provide details about how particular data were generated.

Clarifying how data were generated can help show whether or not the process was influenced by particular biases in information collection and data generation (such as sampling bias). More broadly, rendering data-generation procedures more transparent encourages authors to be more self-conscious about how they collect information and about interpretations of subjective

⁷ Of course, published secondary sources can also be difficult to find or be written in foreign languages.

⁸ When an important subset of the sources cited in a publication (or at least an excerpt therefrom) are made available via the ATI Data Supplement, ATI can also empower the reuse of data. Potential secondary users of these data can immediately examine and download them, and use them for their own scholarly purposes. Obviously, the feasibility and utility of secondary analysis of the data that are shared via ATI will depend on how much information was made available, and to what degree those data constitute a complete and coherent set.

experiences, renders research results more legitimate and tractable, and facilitates scholarly debate and criticism.

ATI and Analytic Transparency

ATI facilitates analytic transparency by providing a clear account of the links between an evidence-based interpretation or descriptive or causal inference in a text and the particular data on which it is based. In doing so, it builds on traditional citation practices, thereby preserving unique epistemic characteristics of qualitative research. ATI improves analytic transparency in three ways.

First, ATI links individual pieces of evidence to aspects of an inferential or interpretive analytic process. In scholarship that draws on qualitative data, those data are typically analyzed, and used to support claims, individually or in small groups. That is, the content of each cited source (e.g., book, archival document, interview transcript, newspaper article, video clip, etc.) serves as a distinct input to the analysis. Moreover, data, analysis, and conclusions are densely interwoven across the span of books and articles, linked to specific stages, aspects, or components of a narrative, analytic process, or causal sequence. By allowing scholars to connect particular pieces of evidence to specific passages in their manuscript, ATI makes transparent the way inferences are being drawn or interpretations are being made. Annotations can be tailored to the characteristics of the qualitative data and the type of inference or interpretation the authors seek to draw or make.

Second, ATI places arguments in a broader evidentiary context. Traditional citations connect claims or conclusion to a reference or, at best, a very short quotation. By contrast, annotations can link an argument advanced in the main text of a publication to an excerpt from the source(s) the author is positing supports the arguments. The excerpts are of sufficient length for a reader to judge whether the source actually supports the interpretation, or descriptive or causal inference, for which it is being called on as evidence. The excerpts offer some assurance that authors have not taken evidence out of context – and providing the sources themselves offers additional assurance. Moreover, providing the text of primary sources such as meeting minutes or interview transcripts offers readers direct access to the voices of the actual subjects of analysis and a richer understanding of the specific cultural, strategic or social context from which particular social facts emerge.

Third, ATI offers researchers an opportunity to “show their work.” In analytic notes, authors can describe how they used the cited data to arrive at the descriptive claim, interpretation, or causal inference discussed in the main text. While extended footnotes can sometimes provide some of this information, traditional citations more typically simply juxtapose claims and references without explaining the precise connection between them. Readers must then engage in “reverse engineering” to identify and evaluate the link being posited. As Moravcsik (2012a, 34) notes, in order to effectively evaluate evidence-based claims, readers need to know “exactly how and why the citation supports the textual point.” Analytic notes permit scholars to document a wider and richer range of descriptive, interpretive, and causal arguments and to underscore their strength; to acknowledge sources of complexity, uncertainty, or ambiguity; to anticipate and answer criticisms; or to clarify methodological issues not fully explained in the text or footnotes. They

can also be used to detail why and how scholars aggregated individual findings to arrive at a particular claim, and how such claims were in turn aggregated to draw conclusions.

V. Elements of an ATI Data Supplement

An ATI Data Supplement is a standardized means to present information about data – how they were generated or collected and how interpretations and inferences were drawn from them – and the data themselves. Creating an ATI Data Supplement assists scholars in efficiently conforming to standards of data access, production transparency and analytic transparency. An ATI Data Supplement’s level of detail depends upon community standards in the author’s research tradition, the particular project and publication, and the researcher’s ambition with regard to rigor and precision. All ATI Data Supplements include two sections: an “ATI Data Overview” and a series of annotations. We discuss each in turn below.

ATI Data Overview

The ATI Data Overview, of approximately 1,000 words in length, serves as an introduction to the ATI Data Supplement. The core goal of the Data Overview is to describe a publication’s empirical base and to offer a holistic, synthetic, overarching discussion of how the data were produced and analyzed – to the degree such discussion cannot be accommodated in the main text.⁹ The Overview thus aids authors in achieving both production and analytic transparency.¹⁰ When publications have a bibliography or list of references, this is also included in the overview.

In discussing how the publication’s evidentiary base was produced, authors should detail the context of data collection and the procedures used to identify, access, collect, generate, capture, prepare, and select data for inclusion as evidence in the publication. Questions of the following type can be addressed: What is the overall field of information that is potentially relevant to the central research question? How did the author decide which information or evidence to consult, to analyze in detail, and to cite?¹¹ What relevant data were not or could not be consulted and what is the potential effect on the analysis? What instruments and procedures were employed to collect data? If interviews, focus, groups, or oral histories were employed, what procedures were used to identify and access potential respondents; what terms of confidentiality and informed consent were employed; and what interview protocols, topic guides, and instructions were issued? Were data photographed, scanned, photocopied, recorded, or summarized in notes? How were data prepared for sharing?

⁹ Authors may also use the Data Overview to make holistic or overarching statements about its research design and logic of inquiry.

¹⁰ In the natural sciences, the evolution of research projects can be traced retrospectively 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 s/he generated and evaluated interpretations, data and inferences, and arrived at conclusions.

¹¹ For example, an author studying a British foreign policy decision might indicate that she 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. She might also justify how the particular private papers, reports, and interview transcripts are relevant to the question and 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 other actors; explain how she interpreted or coded the evidence from these sources; and account for the selection of particular evidence for citation.

Making a publication transparent also requires that authors clearly describe how data were analyzed or interpreted and clearly demonstrate that they have followed the methodological approach they are using (e.g., interpretive analysis, ethnography, grounded theory, process tracing, counterfactual analysis, pattern matching, congruence testing, and *fs*/QCA and other set-theoretic approaches).

Annotations

While the data overview described above is important, the ATI Data Supplement mainly comprises a set of annotations. The annotations are directly linked to the text in the publication. Authors can use most common writing tools (e.g., Microsoft Word) to generate annotations. Detailed instructions are available on the QDR webpage: <https://qdr.syr.edu/deposit/ati-instructions/>. QDR staff convert authors' annotations into the final ATI format, which uses standard web annotations.

As mentioned above, each annotation includes a combination of *one or more* of the following elements (no single element is mandatory):

- **Citation(s)**: full citation for the source(s) mentioned in the annotated portion of the text (when a full citation was not included in the bibliography), and, where relevant, additional location information;
- **Source excerpt**: typically 100 to 150 words from a textual source (such as the transcription of handwritten or audiovisual material);
- **Source excerpt translation**: if the excerpt is not in English, a translation and the source of the translation;
- **Analytic note**: discussion that contextualizes the source(s) and/or illustrates how it was generated and/or analyzed and/or how it supports empirical claims in the text.
- **Data source**: the file name(s) of the corresponding data source(s) linked to the source(s) themselves when these are digital and can be shared ethically and legally.

Citation

A full citation need only be included in an annotation when that information is not already in the text, the footnote/endnote being annotated, or the bibliography of the publication being annotated. When authors wish to annotate several sources in aggregate/holistically (as they may wish to do, for instance, when the sources are all located in the same footnote), all of the citations can be included in one annotation. When authors wish to annotate sources separately, a separate annotation can be created for each.

Source Excerpt

Including brief excerpts (in the original language) from textual or non-textual sources (e.g., an audio clip from a longer recording) referenced in the annotated portion of the publication allows for inferences or interpretations to be understood and evaluated with materials immediately at hand; renders sources accessible and searchable for scholars interested in locating, examining, reusing or downloading source material; and mitigates problems of outdated, inoperative, unstable external links, and unreadable source documents, which bedevil efforts to hyperlink footnotes in other fields.

With regard to the suggested length for source excerpts, we encourage scholars to provide as much relevant evidence as possible. For all sources, textual and non-textual, the excerpt should be sufficient to show the nature and content of the evidence, and for readers to interpret that evidence properly and with nuance. That said, the excerpt should not violate intellectual property or human participant protections, and producing it should not unduly increase logistical burdens for the researcher.

We recommend that excerpts from textual sources be between 100 and 150 words in length.¹² This length is often sufficient to convey the general meaning and context, lies within current scholarly practices with regard to the “fair use” of most types of intellectual property, and imposes a manageable burden with regard to transcribing and de-identifying audio or video recordings.

In some cases, an author may be unable to provide a full verbatim quote of 100-150 words. For example, an author may not possess a source in verbatim form (e.g., no interview notes were taken), it may be lost, or may not be shareable due to human participant, intellectual property, and logistical constraints. In such cases, the author should 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.¹³ While verbatim quotes may be problematic in many research circumstances, there are probably few circumstances in which the author can provide no information about the data s/he used.

Source Excerpt Translation

If the source is not in the language of the publication being annotated, authors should consider providing a translation in the annotation. The name of the translator, date of translation, and other reference information (if any) of the translation should be given. Translation imposes a logistical limit on length; scholars might consider translating just the most crucial portion of the excerpt.

Analytic Note

Analytic notes can discuss the generation and analysis of cited source(s). They can also help to clarify the relationship between those sources and claims, inferences, and interpretations made in a publication by specifying how the cited evidence relates to and supports the passage being annotated. Authors can also use analytic notes to amplify a certain point or to position their claim vis-a-vis that of another scholar.

Analytic notes describe the analytic path between cited data and evidence-based claims. The precise content of a note depends on the type of claim the author seeks to substantiate; the particular principles, logics, or rules of the analytic methods she is employing; the type of data; and the extent of ambiguity or complexity in the evidence. Analytic notes might highlight both confirming and disconfirming evidence, and/or point out subtleties, nuances, contradictions and ambiguities in the evidence. If a textual claim is contested, the note may discuss evidence that *refutes* potential criticisms, perhaps referencing multiple and competing sources not cited in the

¹² ATI guidelines are still being developed for non-textual sources such as visual representations, music, etc.

¹³ Workshop participants with questions or concerns about these constraints should contact QDR.

text. For authors using ATI retroactively, analytic notes might be used to add information that emerged after publication.

Analytic notes may be used to support descriptive claims or causal claims. Analytic notes supporting *descriptive and/or interpretive 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 notes can substantiate a range of descriptive assertions. For example, they may substantiate simple statements of fact (*A* occurred on a particular date). They may also elucidate 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.).

Analytic notes supporting *causal claims* supplement the text by making clear how the evidence shows that the asserted causal relationship holds. The specific content of these notes depends on the particular principles, logics, or rules for drawing causal inferences that guide the analytic methods an author is employing, and the type of data she is using. For example, a note might explain how a particular piece of evidence shows that a causal mechanism was operating in the way the author asserts. Put differently, the persuasiveness of an author's causal claims depends on the degree to which his presentation and discussion of evidence (in the text and annotations) suggests that she has followed the relevant methodological rules.

With regard to length, analytic notes may range from a one-sentence statement 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 the main text, an extremely short note pointing to a critical passage or underscoring a relatively straightforward interpretation may be sufficient.

Data Source(s)

Where possible, authors are encouraged to provide access to full original sources (e.g., audio or video recordings or transcripts from interviews, focus groups, or oral histories as well as translations thereof; ethnographic or participant-observer field notes; archival documents; diaries and other personal records; press clippings; research notes; photographs, maps, posters and other representational work, or artwork). Authors should include the filename of each relevant source in the annotation; data sources will eventually be deposited with QDR.¹⁴

VI. ATI: Standards and Practices

This section offers suggestions concerning when scholars should consider using ATI, which passages or citations in a particular piece of scholarship they might annotate, and what types of annotations they might use.

¹⁴ QDR offers additional information on types of qualitative data (<https://qdr.syr.edu/deposit/data>), on managing qualitative data and preparing them for sharing (<https://qdr.syr.edu/guidance/managing>) and on file naming conventions (<https://qdr.syr.edu/deposit/ati-instructions>).

At What Point in the Research Process Should Scholars Use ATI?

Scholars can annotate a piece of scholarship and produce an ATI Data Supplement contemporaneously with writing a paper, article, or book chapter, or they can do so retrospectively (i.e., drawing on material collected for and employed in already-published work). Using ATI retrospectively can be more difficult. Authors may not have formally recorded their data-selection and data-collection procedures and the sequence in which they employed them, or may not have retained their notes. Moreover, decisions regarding sampling, interpretation, analysis, and other facets of research are often driven by a complex mix of intellectual and pragmatic considerations that may be difficult for authors to reconstruct after the fact.

Authors participating in the first ATI Pilot Working Group workshop, who are retrospectively annotating articles that have already been published, likely did not have annotation in mind when they were conducting their research. Thus, they may not have fully documented these types of choices. We hope authors will do the best that they can, and will note in their logbooks any ways in which engaging in annotation retrospectively made doing so more challenging.

Which Passages Should Be Annotated?

In most cases, only a subset of passages in a published article or book work will be annotated. Annotations are unlikely to be necessary, for instance, when passages and citations reference alternative social science theories, provide methodological guidance, acknowledge previous work on a topic, offer general empirical background, discuss related examples the author does not intend to develop in detail, establish the general importance or policy relevance of a study, and so on.

With respect to portions of articles or books that are more suited to ATI, different scholars will be most comfortable at different points on the continuum stretching from fully annotating all claims in those sections to not annotating any. Scholars could, for instance, annotate all claims that reference anything other than a well-known fact. They might annotate any passages that involve descriptive or causal inference no matter how big or small. We believe that such exhaustive annotation is unnecessary, potentially counter-productive, and almost certainly prohibitively time-consuming.

Which claims and conclusions to annotate is tightly tied to the question of the uses to which ATI can be put. As noted previously, a major goal of the ATI Initiative is to carefully consider, and to develop a clear typology of, those uses. To provide some preliminary guidance, we offer four criteria that authors might employ to determine which passages of their work to annotate. We encourage authors to develop and follow their own strategy or logic for using ATI to increase their work's transparency, keeping in mind opportunity costs and pragmatic limitations.

- 1) **Centrality of the (evidence-based) claim:** is the evidence-based claim a crucial building block to an author's overall argument or thesis?
- 2) **Importance of the data source:** does the claim reference crucial pieces of evidence underlying the author's overall argument or thesis?
- 3) **Contested or controversial nature of the (evidence-based) claim:** is the evidence-based claim 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 it be questioned by members of those communities?¹⁵

- 4) **Contested or controversial nature of the data source:** does the claim reference data sources that are controversial within the research community, debate, or literature to which the author is speaking and on which they draw, or might their validity or evidentiary value be questioned by members of those communities?

Do All Annotations Have to Include Both Excerpts/Sources and Notes?

When annotations are used to amplify an inference or interpretation, we would often expect them to contain an excerpt from the underlying data source (and ideally link to the source itself), as well as an analytic note. Under some circumstances, however, annotations may include an excerpt/source without an analytic note, or vice versa. Annotations with a source/excerpt but no note can be useful when the meaning of the source is unambiguous and uncontested, how it was generated and analyzed is clear, and how it supports the relevant evidence-based claim is patent simply from its juxtaposition with the text and citation, or is clearly stated in the text. Annotations with a note but without an excerpt/source might be used when a cited source documents an unambiguous and uncontested historical fact, but how the source was generated or analyzed, or how it supports a passage in the text, requires elaboration *and* longer citations are not permitted by a publication or space limitations restrict their use.

¹⁵ To offer an example: the word count of the U.S. Declaration of Independence depends on which words one counts: signatures or not, headings or not. While this ambiguity may not matter for most studies, projects concerning sovereignty statements might require source material on these points. Also, scholars can consider annotating 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 (Lustick 1996).

Bibliography

- American Political Science Association. 2012. *A Guide to Professional Ethics in Political Science*. 2nd ed. Washington, DC: American Political Science Association.
<http://www.apsanet.org/portals/54/Files/Publications/APSAEthicsGuide2012.pdf>
- Goertz, Gary, and James Mahoney. 2012. *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences*. Princeton, NJ: Princeton University Press.
<https://doi.org/10.1515/9781400845446>
- King, Gary, Robert O Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.
- Kratz J and Strasser C. 2014. Data publication consensus and controversies [version 3; referees: 3 approved]. *F1000Research*, 3:94 <https://doi.org/10.12688/f1000research.3979.3>
- Lustick, Ian S. 1996. "History, Historiography, and Political Science: Multiple Historical Records and the Problem of Selection Bias." *American Political Science Review* 90 (3):605–18.
<https://doi.org/10.2307/2082612>
- Moravcsik, Andrew. 2010. "Active Citation: A Precondition for Replicable Qualitative Research." *PS: Political Science & Politics* 43 (1):29–35. <https://doi.org/10.1017/S1049096510990781>.
- . 2012a. "Active Citation and Qualitative Political Science." *Qualitative & Multi-Method Research* 10:33–37. http://www.maxwell.syr.edu/uploadedFiles/moynihan/cqrm/Newsletter10_1.pdf
- . 2012b. "How to Create Active Citations Using MS-Word." Princeton, NJ.
http://www.princeton.edu/~amoravcs/library/active_citations.docx
- . 2014a. "Transparency: The Revolution in Qualitative Political Science" *PS* (January).
- . 2014b. "One Norm, Two Standards: Realizing Transparency in Qualitative Political Science," *The Political Methodologist* 22:1 (Fall).
- . 2014c. "Trust, but Verify: The Transparency Revolution and Qualitative International Relations," *Security Studies* 23:4 (2014), pp. 663-688.
- . 2016 "Qualitative Transparency: Pluralistic, Humanistic and Policy Relevant," *International History and Politics APSA Section Newsletter* (Winter 2016).