Annotation for Transparent Inference (ATI): Selecting a Platform for Qualitative Research Based on Individual Sources

Nicholas Weber, University of Washington
Colin Elman, Syracuse University
Diana Kapiszewski, Georgetown University
Carole Palmer, University of Washington

In any given research tradition, scholars believe that if the rules are followed then knowledge of a particular type will be produced. This understanding of social science underpins King, Keohane, and Verba’s (1994) well-known observation that “the content is the method.” A conclusion is a credible and legitimate claim of its type if and only if it was produced in accordance with methodological prescriptions designed to generate knowledge statements of that kind. This meta-standard has a transparency corollary: social scientists working in rule-bound and evidence-based traditions need to show how they know what they know. The less visible the process that produced a conclusion, the less one can see of the conclusion. A sufficiently diminished view of that process undermines the claim.

What an author needs to do to fulfill this transparency obligation differs depending on the nature of the work, the data that were used, and the analyses that were undertaken. For a scholar arriving at a conclusion using a statistical software package to analyze a quantitative dataset, making the claim transparent would include providing the dataset and software commands. In most quantitative work, there are clear and widely-followed conventions on how to distinguish among, and how to locate, (a) the publication (e.g., article), (b) supplemental materials describing the analysis that produced the claims, and (c) the data that were analyzed.

Research transparency is a much newer proposition for qualitative social science. Correspondingly, the requirements and accompanying conventions are not as well developed. Moreover, qualitative research is heterogeneous. Some communities approach research in ways that are analogous to those commonly employed in quantitative traditions, with qualitative data organized in a matrix (dataset) and analyzed in aggregate. For those communities, understandings of transparency will emerge that are roughly comparable to those associate with standard quantitative approaches.

A trickier set of questions arise for social scientists who generate “granular” data from individual sources, and who analyze those data individually or in small groups. Qualitative social science of this type makes rule-bound evidence-based claims and so data and analysis must both be visible. Because the data are not analyzed in aggregate to produce all of the claims in the text, however, new ways have to be developed to associate the claims with the granular data and their analysis.

One very promising solution for achieving transparent inference in this sort of qualitative scholarship involves producing a digital overlay that augments transparency while preserving the unique characteristics of granular qualitative research. The Qualitative Data Repository has been working on a platform for annotation for transparent inference (ATI) for some time, and has made considerable progress, particularly in specifying what information needs to be surfaced for readers to be able to understand and evaluate published claims. With these requirements in mind, this paper will develop a list of functional specifications and a set of criteria for choosing an annotation standard to use as the basis for ATI.
Some of the choices/changes involved in developing a platform for ATI based on the Open Annotation Model include the following:

- The initial focus of annotation for transparent inference will be on how (co)author(s) can mark up their own text, not on crowdsourced annotation.

- Because existing annotation schema may not have had research transparency among their guiding motivations, some revision may be necessary to accommodate the goals mentioned above.

- Annotation for transparent inference often requires excerpting an underlying source (e.g., archival document or interview transcript) and best practices encourage providing such sources where they are available and shareable; consequently a platform for ATI will need to include a mechanism for associating data and excerpt(s) with particular annotations.

- A platform for ATI will need to allow for the aggregation of content (annotations, excerpts, and data) into a stand-alone transparency appendix (TRAX) that can accompany a published article or book.

- If the TRAX is housed in a repository, the platform for ATI will need to provide and support seamless, persistent links between the published article/book and the TRAX.

Having developed a list of criteria for scoring how well candidate annotation schema can satisfy ATI's needs, the paper will provide a preliminary review of several existing annotation platforms.