Transparency in Practice: Working with Annotation for Transparent Inquiry (ATI)

Colin Elman, Diana Kapiszewski, and Sebastian Karcher, QDR
Qualitative Data Repository (QDR): QDR stores, preserves, and publishes digital data arising from qualitative and multi-method inquiry.

QDR is collaborating with Hypothesis, a software non-profit that develops open source software enabling the creation, storage, and sharing of web-standard annotations.
• Annotation for Transparent Inquiry (ATI) is a technique developed by QDR for achieving transparency in qualitative and multi-method research.

• ATI is based on the concept of “open annotation,” which allows for the generation, sharing, and discovery of digital annotations across the web.

• Holding two workshops to demonstrate and evaluate ATI. First involves authors of recently published articles; second authors of manuscripts in progress. Authors are drawn from health and social science disciplines.

• First workshop held, reporting some preliminary results.
QDR helps scholars to share their data to support three missions:

• to enable *secondary analysis*, i.e. so another scholar can analyze the data
• for *pedagogical purposes*
• to achieve *research transparency*
• Evidence-based inquiry is process-dependent, but so are almost all intentionally created outcomes.

• Not just that it is process-dependent, but that the central claim is that “the content is the method.”

• To put it another way, the process is part of the outcome. If you can’t see how the result was arrived at, you can’t see all of the result.

• Principle applies to all rule-bound evidence based inquiry, but application needs to respect the challenges and opportunities that characterize various research traditions.

• Much better understandings and infrastructure for achieving transparency with quantitative than qualitative approaches to research.
Contrasting traditions of social inquiry

- Different ways to organize and typologize research.
- For example, could be by ontology (topic, puzzle, mechanism), discipline, episteme.
- From a transparency perspective, helpful to separate epistemically, by type of data and mode of analysis. In evidence-based inquiry focused on causal analysis, for example, can locate two central groupings: dataset and singular sources.
Quantitative Research: Matrix Data
Qualitative Research: Granular Data

Analysis

Analysis

Analysis

Analysis
Hungary, this was their only stated concern. However, many states conditioned their recognition decision on an action related to Indian troop withdrawal and gave three different types of reasons for doing so. States also differed in the extent of troop withdrawal they required before recognition. See Table 2 for a full list of states, their stated reason for conditioning recognition on withdrawal (if any can be identified), and what recognition was conditioned on (whether actual withdrawal or a proxy).

The first type of reason, opposition to condoning or legitimizing aggression, is labeled as “Non-aggression.” A good example comes from Mexican Foreign Minister Emilio Oscar Rabasa who reported that the Mexican president had decided not to recognize Bangladesh because, “since the Mexicans, like many Latin Americans, refuse to condone territorial aggrandizement as a result of war, they would prefer to wait on the withdrawal of Indian troops as the sign of true independence.”

This statement also appeals to “true independence.” Self-determination is another important value expressed by the Mexican representative and is the second type of reason commonly appealed to as justifying recognition as Bangladesh. For

88. See Figure 2.
89. A frequent concern was that states had to recognize in a group, or on the same day as multiple other states. However, even allowing for minor coordination problems, this in and of itself cannot explain the length of time taken to make recognition decisions and declarations.
90. Cable from Hope, 16 January 1972, FCO 371/020.

ATI Annotation: Displayed alongside article. Created by author, curated by QDR, hosted and served by Hypothesis, displayed on publisher’s website

Elements of an ATI annotation:
- Analytic Note
- Source Excerpt
- Source Excerpt Translation
- Link to Data Source

Any digitally published manuscript can be annotated using ATI (here: an article in International Organization published by Cambridge University Press)

Any passage in the text or in notes of a manuscript can be annotated using ATI

Link to data source housed in QDR
Two ATI examples


ATI Initiative

• Includes the ATI Pilot Working Group
• Authors from various health and social science disciplines who employ multiple analytic approaches, use diverse types of qualitative data, and study different world regions and temporal periods.
• Reviewers, charged with evaluating authors’ use of ATI according to a pre-determined set of criteria.
The ATI Pilot working group by the numbers:

37 Scholars (20 authors, 17 reviewers)

17 Articles

7 disciplines (Anthropology, Health, Public Health, Linguistics, Political Science, Psychology, Sociology)

379 Annotations

159 Data source files
Some of the issues the ATI working group is trying to address

• Does ATI help make qualitative research more transparent?
• For what articles, methods, disciplines does it work particularly well or particularly poorly?
• Does ATI have benefits/uses other than transparency?
• Could the structure of ATI be improved?
• How well is the technology we use suited for the purpose and how could it be improved?
• How can the documentation be improved?
Learning about ATI from the working group

• Authors
  ▪ Were asked to keep logbook while annotating
  ▪ Were asked to complete questionnaire

• Reviewers
  ▪ Were asked to read/assess/evaluate original (i.e. pre-annotation) article, and compare with annotated version.

• Workshop
  ▪ Thematic panels plus group discussion of all 17 annotated articles
Logbook questions for the annotating authors

We asked authors to keep a logbook while they annotated, answering questions that included:

1. Why are you choosing particular passages, citations, or footnotes of the manuscript for annotation?
   a. Are you annotating particular types of claims (e.g., descriptive, or causal, or controversial)?
   b. Do you find you’re drawn to annotating claims in some sections of your article more than others?

2. Where you are including analytic notes, what is their main function (e.g., do they offer additional context, reflections, or interpretation; evaluate sources; discuss how sources were produced or analyzed; or elucidate links between evidence and claims)?
Logbook questions for the annotating authors

3. Why are you choosing to link underlying data sources (i.e., deposit a file with QDR and link to it in the annotation) to particular passages, citations, or footnotes of the manuscript? Where you choose not to link in underlying data sources, what are some of the reasons?

4. How did the way you organized your notes and files impact your ability to annotate your article?

5. Are there any aspects of our instructions for creating an ATI Data Supplement that are unclear, or any processes that can be simplified?
Tracking down data sources

• How hard/easy was it to find data sources on which claims were based on computer/in records?

• Based on retrospective annotation of a previously (albeit, recently) published article. May look different for contemporaneous annotation.
Difficulty Finding Data Source

Reported difficulty of finding data sources

Number of respondents

Very easy
Easy
Moderate
Hard
Very hard
Not applicable
Tracking down *origin* of data sources

- How hard/easy was it to find the origin of the data sources on which claims were based?
- Again, based on retrospective annotation of a previously (albeit, recently) published article. May look different for contemporaneous annotation.
Difficulty Finding Data Source Origins

Reported difficulty of finding data source origins

Number of respondents

Very easy | Easy | Moderate | Hard | Very hard | Not applicable

0 | 5 | 10 | 15 | 15 |
Determining source rights (copyright and/or license) of data sources:

• How hard/easy was it to determine whether the data sources were subject to proprietary constraint?
Difficulty Finding Source Rights

Reported difficulty of finding data source rights

Number of respondents

- Very easy
- Easy
- Moderate
- Hard
- Very hard
- Not applicable
Determining whether impact of human subjects protections?

- How hard/easy was it to determine whether guarantees offered to human participants affected sharing data?
Difficulty Identifying Participant Protections

Reported difficulty of identifying protections of human participants

Number of respondents

- Very easy
- Easy
- Moderate
- Hard
- Very hard
- Not applicable
Preparing data for publication?

• How hard/easy was it to prepare data for archiving (e.g. converting paper to digital, generating a pdf, scanning a document, etc.)?
Difficulty preparing data for publication
Choosing what to excerpt in annotations?

• How hard/easy was it to choose relevant excerpts of data sources to include in annotations?
Difficulty of Choosing Excerpts

Reported difficulty of choosing excerpts

- Very easy
- Easy
- Moderate
- Hard
- Very hard
- Not applicable

Number of respondents
Translating excerpts?

• How hard/easy was it to translate excerpts for the annotations?
• Only relevant for non-English sources, and optional.
Difficulty translating excerpts

Reported difficulty of translating excerpts

Number of respondents

- Very easy
- Easy
- Moderate
- Hard
- Very hard
- Not applicable
Creating analytic notes about data generation?

• How hard/easy was it to create notes about how data sources were generated?
Difficulty writing notes on data generation
Creating analytic notes about data analysis?

• How hard/easy was it to create notes about how data sources were analyzed?
Difficulty writing notes on data analysis
## Distribution of answers across articles

<table>
<thead>
<tr>
<th>Question</th>
<th>A</th>
<th>N</th>
<th>Q2</th>
<th>J</th>
<th>K</th>
<th>H2</th>
<th>G</th>
<th>B</th>
<th>O</th>
<th>P</th>
<th>F</th>
<th>Q1</th>
<th>L1</th>
<th>L2</th>
<th>C</th>
<th>E</th>
<th>H1</th>
<th>D</th>
<th>I</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Finding data sources</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>2. Origin of data sources</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>3. Source rights</td>
<td>2</td>
<td>3</td>
<td>na</td>
<td>1</td>
<td>na</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>na</td>
<td>1</td>
<td>2</td>
<td>na</td>
<td>na</td>
<td>1</td>
<td>na</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>na</td>
<td>1</td>
</tr>
<tr>
<td>4. Human participants afl</td>
<td>5</td>
<td>na</td>
<td>na</td>
<td>3</td>
<td>na</td>
<td>1</td>
<td>na</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>na</td>
<td>na</td>
<td>2</td>
<td>1</td>
<td>na</td>
<td>na</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>5. Preparing data</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>na</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>na</td>
<td>2</td>
<td>3</td>
<td>na</td>
<td>3</td>
<td>3</td>
<td>na</td>
<td>na</td>
<td>2</td>
</tr>
<tr>
<td>6. Choosing excerpts</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>7. Translating excerpts</td>
<td>1</td>
<td>na</td>
<td>3</td>
<td>2</td>
<td>na</td>
<td>3</td>
<td>na</td>
<td>3</td>
<td>3</td>
<td>na</td>
<td>3</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>3</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td>na</td>
<td></td>
</tr>
<tr>
<td>8. Data Notes</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>na</td>
<td>3</td>
<td>3</td>
<td>na</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>9. Analysis Notes</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>na</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>
After the workshop

- ATI projects from February workshop on articles published by Cambridge University Press are live.
- Second group of ATI projects from articles from non-CUP journals will launch soon (waiting on some minor teach tweaks from Hypothesis).
- ATI Challenge launched, meritocratic selection of ATI projects for second (November) workshop. Difference from first workshop, annotate contemporaneously and not retrospectively.