How Insurgent Artifacts meets the DDI standards of metadata

The data documentation initiative (DDI) standards are a set of metadata and data collection standards and tools for the social sciences that have been developed over the last 25 years (<https://ddialliance.org/what/history.html>). They provide a systematic and computer searchable form of recording data and metadata. Though originally developed with survey questionnaire data collection and metadata management in mind, they have been expanded in recent years to accommodate geographic mapping projects and measurement-based projects. All of these applications rely on the ability to consistently capture similar types of information in an XML database. Qualitative, semi-structured interviews and observations do not have the same degree of structure to their collection as these other applications and so the DDI platform does not directly apply as a tool for metadata and data collection archiving. Instead, the Insurgent Artifacts project based its data collection, coding, and metadata documentation around the principles that guide the DDI data lifecycle (see Figure 1, below).



Figure 1. DDI data lifecycle (source: <https://ddialliance.org/training/why-use-ddi>)

To the extent possible, the Insurgent Artifacts used this model to inform how we developed and managed the metadata, data collection, and data coding associated with the project. Below is a diagram of how our metadata process mirrors the DDI life cycle, where applicable (see Figure 2). We prioritized documenting the decision-making process at each point in Figure 2 to create metadata at each point, from conceptualizing the research goals of the project to conducting analysis of the data. These processes were captured in word documents, codebooks, and emails between the PI and postdoctoral fellow. All formal decisions were recorded in a “metadata” folder in the project filing system so that they were easily accessible. The filing system incorporated version control so that updates to these files were automatically captured and archived. A key example of how the project mirrored the DDI process is in the data coding and analysis part of the process. DDI 3.0 requires creating variables by defining categories, assembling them into category coding structures and then using these structures to create variables (see <https://ddialliance.org/getting-started>). To code the interviews, we used the research questions for the project to define broad categories and then create a coding structure for key terms that were then used to operationalize key variables. For example, identifying where and how interviewees conceptualized “standards of evidence” in their work was an important variable for the analysis and was interlinked conceptually with how interviewees viewed “credibility and reliability.” To capture the nuances in these variables, we developed a coding structure and codebook that defined the uses for each code. Standards of evidence was captured by a code “standard” that was clearly defined to distinguish it from instances where it may overlap with credibility. If an interviewee responded to a question about credibility and reliability with a detailed description of their process, then that was a signal that they were describing their standard for what constituted evidence and was coded as “standard” instead of “credibility.”

For archival data, set protocols for collection were developed up front based on the research goals and updated as appropriate throughout the process. All updates were archived in the filing system. Coding for archival research was done using systematic codebooks that formalized key terms and the concepts in question. All forms of data coding incorporated interrater reliability testing to further document the process of creating the final datasets.

ID Research Goals

Data Collection

Data Coding

Review

Analysis

Discovery

Distribution

Metadata Documentation

Archiving

Figure 2. Insurgent Artifacts Data Lifecycle. Dashed lines indicate an the flow of metadata and archiving.