Memo for qualitative coding

Goal: hybrid approach for developing a coding scheme (first deductive, then inductive, then deductive again)?

Step 1a): I create a tabular check list with indicators of theoretical components. The information from that table informs the creation of NPM-related and DEG-related terms for the dictionary that consists of some of the 50 most frequent words used in all texts (including all stakeholders’ statements and relevant EU documents)

Step 1b): Using the text mining software, R Studio, I identify the 50 most frequently used words in the texts by ranking them in the descending order of frequency visually. Next, I conduct a cluster analysis, via (circular) dendrograms, to identify potential word combinations of 2-3 words that may align with the description of the theoretical components. I create a dictionary that consists of the word combinations that are categorised into one of the theoretical components.

Step 1c): That dictionary of NPM-related and DEG-related terms guides the process of phrase matching to put the terms into context. To learn more about the context of the key words created via the word clusters, I extract the statements related to one of the word clusters, using the quanteda package of R Studio. (Reason: limits of quantitative text mining/extraction)

Step 2a): Looking at these statements by each word cluster, I kick off the first coding cycle with open coding by identifying preliminary codes for these statements within a given context, keeping the hypotheses and the main research question in mind. Only statements that provide a potential answer to the research question are coded.

Step 2b): In preparation for the axial coding stage, I consolidate the codes into groups based on their commonalities, because this is a reiterative process.

- Initially, each cluster has a set of codes, the length of which varies depending on the number of search results I found via the text mining tool in R by extracting the relevant statements for the hypotheses and research question.

- Across the clusters, I merge the codes based on their commonality in terms of their potential applicability to one of theoretical components.This is a reiterative process.

- The dictionary is only helpful to a limited extent: Before conducting open coding, I assigned each word combination (identified via the cluster analysis in R Studio) to a theoretical component based on the latter’s indicator defined earlier. During the open coding stage, I realised that some clusters do not necessarily match their prescribed theoretical component in the statements that contain the word combined but match another theoretical component (given that a theory applies at all). Depending on the context of the statement, the meaning of a term can vary.

- I identify preliminary axial codes by categorising the codes into the same theoretical components first. To this end, each code is assigned a distinct theoretical component; some components have more than one code. As there are some single codes that match one theoretical component, however, I organise the final axial code along the theoretical theme. There are six themes in total, three for each theory, but none of the data match any of the components of disaggregation, an NPM theme. As a result, I label the axial codes based on the commonality between the theoretical components assigned to these codes in accordance with the corresponding theoretical theme.

Step 2c): Next, I move to the axial coding stage, in which I match the codes to the theoretical components of NPM or DEG (given that one of the theories applies). This is when the dictionary as well as the check list come into play because they inform the process of finding matches between the theoretical components and the statements (see Step 1). As qualitative coding represents a reiterative process, the coding process ends, within the context of this study, once there are no concepts can be further extracted from the qualitative data.