

Identification of Topics Overlapping with Theme „The Actual or Desirable Public Role of Religion in Societal Life”

Selection Protocol (v2.1)

When selecting topics as relevant or not relevant, the following selection protocol applies:

1. Interpretability-/coherence rule: Does the topic (i.e., the word combination that characterizes it) convey a coherent meaning? Note: The meaning may be loose but must be clearly recognizable.
 - If no → do not mark as relevant topic, continue with step 3
 - If yes → continue with step 2

2. Relevance rule: Is it plausible that the topic contains any articles that are thematically relevant to our project keeping our document selection protocol in mind?
 - If no → do not mark as relevant topic, continue with step 3
 - If yes → mark as relevant topic, continue as step 3

Note 1: It can also be plausible that a topic contains thematically relevant articles if the words that make up the topic do not explicitly refer to religion (for example, a topic dealing with “homosexuality” would be thematically relevant based on our prior knowledge even if it does not contain religion-related words).

Note 2: Independent of your own certainty or uncertainty as to whether a topic is thematically relevant (see rule 3), every topic should be coded without exception; if you are uncertain about how to code a topic, apply the code you are leaning to more (tendency coding).

3. Case-of-doubt rule: In addition to coding the relevance, note for **each topic** whether you were certain or uncertain about your code. There is a separate column to take note of your own certainty or uncertainty while coding the topic’s relevance.

Important general note: In principle, the identification of the relevant topics should be **generous / liberal**. Since the articles, that are automatically selected on the basis of the identified topics, are manually checked for their thematic relevance in a follow-up step with the Online Relevance Coder, the most important goal here is to sort out all the topics that clearly do not make sense and that are highly unlikely to contain relevant articles. Topics, that could at least partly contain relevant articles, should be included, even if they are likely to also include irrelevant articles.

4. Coding of topic coherence: In addition to steps 1-3, note for **each topic** how coherent the topic is, that is, how strongly the words of the topic are substantively related. There is a separate column for noting each topic's coherence. In this column, please note for each topic how coherent it is (keeping the rule below in mind) on a 3-point scale from **1 = largely incoherent to 3 = largely coherent**. In order to assess the coherence of each topic **consider the top-10 words of the topic** ("Term 1" to "Term 10").

A topic is coherent if its top words have a recognizable context of meaning; if the topic can easily be given a title (i.e., a "label") and if it can be interpreted and used as a thematic heading. For our purposes, topic coherence means that you can imagine using the topic (i.e., its top words) in a search mask to find documents about a particular topic. One indicator of coherence is the ease with which you can think of a short title to describe the topic.

Example incoherent topic (code "1"): stories, undated, receive, scheduled, clients, running, basket, George, tower, Quran

Example coherent topic (code "3"): farmers, farm, food, rice, agriculture, crop, harvest, ranchman, wheat, soil

Topics should be assigned code "2", if they are **neither largely incoherent nor largely coherent in the above sense**.