Data Documentation

Every single word – A new dataset including all parliamentary materials published in Germany

Data description:

The dataset includes all parliamentary materials published by the German Bundestag between 1949 and 2017 (except debates).

In the main dataset, each document constitutes one observation. We additional provide a dataset including separated information for requests, given that all requests submitted for a plenary meeting are published in a single document. In this second dataset, each request constitutes one observation.

The variables include (1) meta data describing the document, (2) author-related information, and (3) the full text.

Variables in the main dataset:

Variable name	Description
number	Running number [Drucksache].
document_type	Original description of document type
leg_period	Legislative period, 1-18.
date	Date at which the document was submitted.
author_i_001 – author_i_009	Name of author (institutions); the dataset includes documents jointly authored by up to nine distinct institutions.
author_mp_001 – author_mp_448	Name of author (MPs); the dataset includes documents jointly authored by up to 448 MPs.
author_mp_id_001 – author_mp_id_448	MP ID for Bergmann et al. (2018).
text	Plain text without meta and author information as well as headers.

Variables in the separated request dataset:

Variable name	Description
number	Running number [Drucksache].
document_type	Original description of document type
leg_period	Legislative period, 1-18.
date	Date at which the document was submitted.
ministry	Name of the addressed ministry
author	Name of author (MPs).
mp_id	MP ID for Bergmann et al. (2018).
question	Plain text of the request.