

The EPINetz Twitter Politicians Dataset

For the initial data acquisition, we built on a number of Twitter lists of German politicians in parliament curated and made publicly available by Martin Fuchs aka “Wahlbeobachter” (election observer).¹ In order to assess this data, we relied on an early, non-public version of the Social Media Observatory (SMO)’s Twitter Parliamentarian Database (Münch et al. 2021).² While overlaps between the two data sources were included into the EPINetz Twitter Politicians Dataset right away, discrepancies were hand-evaluated before including them. After the initial evaluation in early 2021, for every state or federal election thereafter, we scanned the official parliament databases for new members elected and searched for these members’ Twitter accounts by hand. At the same time, we tracked which politicians in our dataset were still in office and noted their respective terms of office. Note that, due to the initial evaluation taking place in early 2021, office terms are only tracked for those legislative periods stretching into and starting in 2021. See Table 1 for a comprehensive overview of legislative periods covered. Finally, in late 2021, after federal elections and numerous state elections were held, we checked back the data so far collected by comparing it with the by-now updated Twitter lists of German politicians in Parliament curated by Martin Fuchs. Again, we hand-evaluated those accounts not already covered in our data before adding them to the database. This list of parliamentarians was then supplemented by an in-depth investigation into the Twitter accounts of German ministers, state secretaries, ministries, and parties on both federal and state level. Due to the multinational character of its institutions, we only collected data for German MEPs on the EU level. Any inconsistencies between data sources and the qualitative search that appeared during re-evaluation were resolved by consecutive lookups using the following resources : 1) Official sources (such as the state website) were consulted to retrieve information on who was part of the current parliament or government. For parties and ministries, this step was skipped. 2) Twitter’s own search option was used to look for accounts with the same name as the parliamentarian, state secretary or minister in question. For parties, we looked for accounts representing the state or federal chapters of all parties elected to the respective state or federal parliaments. For ministries, we looked for all accounts representing ministries on a state or federal level. 3) We checked the account’s description, timelines and displayed name to determine whether the account represented the desired politician or institution before adding it to the database. In some cases, both a private and an official account were available, in which case we opted for the official account. If only a private, but clearly identifiable, account was available, this was also added to the database. Accounts that were inactive and, in some cases, superseded by another account, were also added to the database for archival reasons, e.g., when research interest requires retrieving tweets for a longer period of time.³ In order to obtain additional socio-demographic variables, we used the Wikidata API.⁴ Specifically, we retrieved information on gender and year of birth for all individual persons, and the Abgeordnetenwatch (Parliamentwatch) ID

¹ <http://www.hamburger-wahlbeobachter.de/> / https://twitter.com/wahl_beobachter

² Special thanks to Felix-Victor Münch and the team at the Social Media Observatory for granting us access to this work in progress database.

³ Note that this produces multiple account entries for politicians with multiple accounts, e.g., for the new Foreign Minister, Annalena Baerbock, whose old Twitter account was superseded and renamed into “@ABaerbockArchiv” when assuming office. In these cases, both the old and the new account are available in our database.

⁴ <https://www.wikidata.org/>. Note that we also use the API to supplement the existing entries with the social media information here collected, allowing researchers to access up-to-date information on German politicians’ Twitter accounts directly via Wikidata.

for parliamentarians. When these variables were unavailable on Wikidata, we manually extracted this information from the Abgeordnetenwatch website instead.⁵

Table 1 Legislative Periods Covered

Region	Legislative Period I		Legislative Period II (if available)	
	Parliament	Government	Parliament	Government
Federal (Bund)	2017-10-24	2018-03-14	2021-10-26	2021-12-08
European Union	2019-07-02			
Baden-Württemberg	2016-03-13	2016-05-12	2021-05-11	2021-05-12
Bavaria	2018-11-05	2018-11-12		
Berlin	2016-10-27	2016-12-08	2021-11-04	2021-12-21
Brandenburg	2019-09-25	2019-11-20		
Bremen	2019-06-08	2019-08-15		
Hamburg	2020-03-18	2020-06-10		
Hesse	2019-01-18	2019-01-18		
Lower Saxony	2017-11-14	2017-11-22		
Mecklenburg-West Pomerania	2016-10-04	2017-07-04	2021-10-26	2021-11-15
North Rhine-Westphalia	2017-06-01	2017-06-30		
Rhineland-Palatinate	2016-03-13	2016-05-18	2021-05-18	2021-05-18
Saarland	2017-04-25	2018-03-01		
Saxony	2019-10-01	2019-12-20		
Saxony-Anhalt	2016-04-12	2016-04-25	2021-07-06	2021-11-16
Schleswig-Holstein	2017-06-06	2017-06-28		
Thuringia	2019-11-26	2020-03-04		

Variables in the EPINetz Twitter Politicians Dataset

The EPINetz Twitter Politicians Dataset comes with a number of variables, allowing for a more fine-grained analysis. See Table 2 for a comprehensive overview. The ID represents a unique identifier generated over an accounts' user_id, region, institution, office and party membership. This makes sure the same Twitter account associated with different offices (e.g., a politician is both a member of parliament and a minister), regions (e.g., changing or multiple parliamentary memberships), institutions (e.g., a change from the state to the federal government), or parties (e.g. changes in party membership) can be uniquely identified in their different roles. The official_name represents a person's or institution's name as it is referred to in official documents such as parliamentary lists (and usually the media). As these documents can have inconsistent naming conventions (e.g. between legislative periods), the middle names of politicians are dropped in order to ensure maximum compatibility with other official documents. The From Variable indicates the day a person assumes office. For parliamentarians, this is the first session of the newly elected parliament after elections. For ministers and state secretaries, this is the date of formation of the new government after elections. As mentioned before, we only cover legislative periods lasting until or starting in 2021. Therefore, the earliest possible From date is the beginning of these legislative or governance periods (see Table 1) and longer incumbencies are not covered. The Until variable marks the date a person left office, e.g. when not re-elected into parliament or giving up their ministerial position. The date of reference here is the first session of the newly elected government after election or the formation of the new government. Again, this is only covered for the legislative periods connected to 2021. When a person is currently holding an office, the Until value is set to missing (NA), as the required dates of reference

⁵ <https://www.abgeordnetenwatch.de/>

for upcoming legislative periods are not yet set. Note that we do not comprehensively track changes in parliament in between these dates of reference. Therefore, while changes in governments (e.g., changes in ministerial positions during legislatures) are represented in our data, we only collected data on who was elected into parliament, not on the periods of time they were actually active in parliament. As parties and ministries are not bound to legislative or governance periods, both their From and Until variables are left with empty (NA) values. The Party variable expresses an account's party affiliation. This is left blank for ministries, as they are in itself neutral institutions under changing leadership. This variable can produce duplicates in the data, e.g., when a politician changes parties between legislative periods.⁶ Note that the same dates of reference apply as for the From and Until variables and changes in between legislative periods are not tracked.

The Region variable represents a person's, ministry's, or party's association with a polity level (state, federal, EU). It can be one of the 16 federal states of Germany, "Federal" (Bund) or "EU" for members of the European Parliament. Party accounts associated with a certain state branch of a party are assigned the respective region, while for party accounts without such clear regional indication nationwide activity is assumed and they are labelled as federal. The variable Institution represents an account's institutional association. For parliamentarians, this is the state, federal or European Parliament, while for ministers, state secretaries and ministries, this is the state or federal government. For accounts representing parties, this is labelled as "Party". For these accounts, the Office variable further differentiates between Speakers and Parliamentary Party Groups. While the former indicates a party's official spokesperson or account, the latter indicates accounts specifically associated with a party's voting bloc in a given parliament, which does not necessarily represent all members of a party. For all other accounts, the Office variable indicates the political office, which can either be that of a parliamentarian, minister, state secretary or ministry. Note that the Region, Institution and Office variables are prone to producing duplicates in the data, for example when multiple offices are held, or institutional affiliations change between legislatures. This, however, is a feature, rather than a drawback, of the dataset. The high granularity of data allows to clearly differentiate between different polity levels, roles and offices in any analyses. **It is advised that researchers filter for unique User_IDs before pulling any Twitter data to avoid skewing their data with unintended tweet duplicates.** Afterwards, one can add the additional variables from the EPINetz Politicians Dataset and filter the data according to research interest.

The variables Twitter_name, Screen_name and User_ID represent an account's Twitter presence. While the Twitter_name refers to the name a user choses to display on their Twitter profile, the Screen_name represents their Twitter handle (@screen_name). Both of these variables can be changed at will by the user. Therefore, they are non-persistent and the values in the dataset represent the names and handles as of time of writing. The User_ID represents an accounts numerical representation on the platform Twitter. Note that, unlike the Twitter_name and Screen_name, it cannot be changed by the user and is unique and persistent for all accounts. Therefore, **it is strongly advised that researchers utilise the User_ID when retrieving data via the Twitter API.** The Year of Birth and Gender variables are sourced from Wikidata or, when unavailable, collected from Abgeordnetenwatch (see above). While available for most individual persons (parliamentarians, ministers, state secretaries), they are intentionally left blank (NA) for accounts associated with parties and ministries. The Abgeordnetenwatch_ID represents a person's unique identifier on the independent, non-partisan platform Abgeordnetenwatch.⁷ Abgeordnetenwatch (Parliamentwatch)

⁶ For the case of two parties forming a single parliamentary party group represented by a singular Twitter account, said account is also duplicated with different party affiliations to represent both parties equally. In our data, this was the case only for the CDU/CSU Bundestag faction and its two associated accounts.

⁷ <https://www.abgeordnetenwatch.de/>

tracks incumbencies, candidacies, and voting records for all German parliamentarians. It also allows citizens to publicly pose questions to their representatives through the site. The identifier was included to enable further research avenues, e.g. when trying to associate Twitter activity with voting behaviour or responsiveness to citizens' questions. When unavailable (e.g. for parties and ministries), it was left blank (NA). Finally, we included the Wikidata_ID so researchers could easily collect additional information on the accounts included in the dataset. Again, this variable was left blank (NA) when unavailable.

Table 2 Variables in the EPINetz Twitter Politicians Dataset

Variable	Description
ID	A unique identifier, generated over user_id, region, institution, office and party membership.
Official_name	The official name of a person or institution, as referred to in official documents or the press. Middle names and titles are dropped.
From	Date of a person's assumption of office. Only goes back to the last legislative period. NA for parties and ministries.
Until	Date of a person's dismissal of office. NA for ongoing offices. Always NA for parties and ministries.
Party	Party affiliation of a person or official party account. NA for ministries.
Region	The political region a person's, party's or ministry's activities are focused on. One of the German federal states, the federal level (Bund) or EU.
Institution	The institutional affiliation of an account. State, federal, or European parliament for parliamentarians; state or federal government for ministers, ministries, and state secretaries; "Party" for official party accounts.
Office	Political office associated with an account. Refers to their role as parliamentarian, minister, state secretary or ministry. Accounts associated with parties are split into "Parliamentary Party Group" (for accounts representing parliamentary factions) and "Speaker" (for accounts representing party speakers).
Twitter_name	An account's name as displayed on Twitter. Subject to change by users.
Twitter_handle	An account's twitter handle (@handle). Subject to change by users.
User_ID	An account's unique twitter ID. Cannot be changed by users.
Year_of_birth	A person's year of birth, if available. Always NA for parties and ministries.
Gender	A person's self-ascribed gender, if available. Always NA for parties and ministries.
Abgeordnetenwatch_ID	A person's ID on the platform Abgeordnetenwatch (Parliamentwatch), if available. Only available for parliamentarians and former parliamentarians.
Wikidata_ID	The ID of the associated Wikidata page, if available.