

# Package ‘legislatoR’

October 13, 2022

**Title** Interface to the Comparative Legislators Database

**Description** Facilitates access to the Comparative Legislators Database (CLD). The CLD includes political, sociodemographic, career, online presence, public attention, and visual information for over 45,000 contemporary and historical politicians from ten countries.

**Version** 1.0

**URL** <https://github.com/saschagobel/legislatoR>

**BugReports** <https://github.com/saschagobel/legislatoR/issues>

**License** GPL-3

**LazyData** true

**Depends** R (>= 3.4.0)

**Imports** curl (>= 3.0), dplyr (>= 0.7.4)

**Suggests** testthat, rmarkdown, knitr, magrittr, purrr, stringr, tibble

**VignetteBuilder** knitr

**Encoding** UTF-8

**RoxygenNote** 6.1.1

**NeedsCompilation** no

**Author** Sascha Goebel [aut, cre] (<<https://orcid.org/0000-0002-9032-5874>>),  
Simon Munzert [aut]

**Maintainer** Sascha Goebel <[sascha.goebel@uni-konstanz.de](mailto:sascha.goebel@uni-konstanz.de)>

**Repository** CRAN

**Date/Publication** 2020-04-24 11:30:02 UTC

## R topics documented:

cld_content . . . . .	2
get_core . . . . .	3
get_history . . . . .	4
get_ids . . . . .	5
get_office . . . . .	7

get_political . . . . .	8
get_portrait . . . . .	9
get_profession . . . . .	10
get_social . . . . .	12
get_traffic . . . . .	13
legislatoR . . . . .	14
<b>Index</b>	<b>15</b>

---

cld_content	<i>List content of the CLD</i>
-------------	--------------------------------

---

## Description

Returns a named list of legislatures and sessions available in the CLD. This provides a quick overview of the CLD's scope and valid three-letter country codes, and helps to conveniently loop/map over legislatures and sessions.

## Usage

```
cld_content(legislature = NULL)
```

## Arguments

**legislature** An optional character string specifying one or more legislatures. Currently one of 'aut', 'can', 'cze', 'esp', 'fra', 'deu', 'irl', 'sco', 'gbr', 'usa\_house', or 'usa\_senate'. If NULL (the default), a list with all legislatures and sessions available in the CLD is returned.

## Value

A list with names being three-letter country codes and with each element containing a vector that shows the sessions available for a legislature.

## Examples

```
# Get a list of three-letter country codes and available sessions for all countries
overview <- cld_content()
tibble::glimpse(overview)

# Get a list of available sessions for the French Assemblée and the Irish Dail
sessions <- cld_content(legislature = c("fra", "irl"))
tibble::glimpse(sessions)
```

---

`get_core`*Fetch 'Core' table*

---

**Description**

Fetches sociodemographic data of legislators for the specified legislature. Requires a working Internet connection.

**Usage**

```
get_core(legislature)
```

**Arguments**

`legislature` A character string specifying the three-letter country code of the legislature for which data shall be fetched. Currently one of 'aut', 'can', 'cze', 'esp', 'fra', 'deu', 'irl', 'sco', 'gbr', 'usa\_house', or 'usa\_senate'.

**Format**

Data frame with columns (varies by legislature):

- `country`: ISO 3166-1 alpha-3 three-letter country code (of class 'character').
- `pageid`: Wikipedia page ID identifying a legislator's Wikipedia biography (of class 'integer' or 'character').
- `wikidataid`: Wikidata ID identifying a legislator's Wikidata entry (of class 'character').
- `wikititle`: A legislator's undirected Wikipedia title (of class 'character').
- `name`: A legislator's full name (of class 'character').
- `sex`: A legislator's sex (of class 'character').
- `ethnicity`: A legislator's ethnicity (of class 'character').
- `religion`: A legislator's religious denomination (of class 'character').
- `birth`: A legislator's date of birth (of class 'POSIXct').
- `death`: A legislator's date of death (of class 'POSIXct').
- `birthplace`: Comma separated latitude and longitude of a legislator's place of birth (of class 'character').
- `deathplace`: Comma separated latitude and longitude of a legislator's place of death (of class 'character').

**Value**

A data frame with columns as specified above.

**Source**

Wikipedia, <https://wikipedia.org/>  
 Wikipedia API, <https://wikipedia.org/w/api.php>  
 Wikidata API, <https://www.wikidata.org/>  
 Wikimedia Commons, <https://commons.wikimedia.org/>  
 Face++ Cognitive Services API, <https://www.faceplusplus.com/>  
 Germany Bundestag Open Data, <https://www.bundestag.de/service/opendata>

**Examples**

```
# Get entire 'Core' table for the German Bundestag
deu_core <- get_core(legislature = "deu")
tibble::glimpse(deu_core)

# Get 'Core' table for 16th session of the German Bundestag
deu_core_subset <- dplyr::semi_join(x = deu_core,
                                   y = dplyr::filter(get_political(legislature = "deu"),
                                                       session == 16),
                                   by = "pageid")
tibble::glimpse(deu_core_subset)
```

---

 get\_history

---

*Fetch Wikipedia 'History' table*


---

**Description**

Fetches full revision histories of legislators' Wikipedia biographies for the specified legislature. Requires a working Internet connection.

**Usage**

```
get_history(legislature)
```

**Arguments**

**legislature** A character string specifying the three-letter country code of the legislature for which data shall be fetched. Currently one of 'aut', 'can', 'cze', 'esp', 'fra', 'deu', 'irl', 'sco', 'gbr', 'usa\_house', or 'usa\_senate'.

**Format**

Data frame with columns:

- pageid: Wikipedia page ID identifying a legislator's Wikipedia biography (of class 'integer').
- revid: Wikipedia edit ID (of class 'integer').
- parentid: Wikipedia edit ID of the previous revision (of class 'integer').

- user: Username of registered user responsible for the revision, IP address in case of anonymous revision (of class 'character').
- userid: ID of registered user responsible for the revision, 0 in case of anonymous revision (of class 'integer').
- timestamp: Date and time of the revision (of class 'POSIXct').
- size: Revision size in bytes (of class 'integer').
- comment: Revision comment (of class 'character').

### Value

A data frame with columns as specified above.

### Source

Wikipedia API, <https://wikipedia.org/w/api.php>

### Examples

```
# Get entire 'History' table for the Austrian Nationalrat
aut_history <- get_history(legislature = "aut")
tibble::glimpse(aut_history)

# Get 'History' table for NEOS party members of the Austrian Nationalrat
aut_history_subset <- dplyr::semi_join(x = aut_history,
                                     y = dplyr::filter(get_political(legislature = "aut"),
                                                         party == "NEOS"),
                                     by = "pageid")
tibble::glimpse(aut_history_subset)
```

---

get\_ids

*Fetch 'IDs' table*

---

### Description

Fetches a range of IDs of legislators for the specified legislature. Requires a working Internet connection.

### Usage

```
get_ids(legislature)
```

### Arguments

legislature      A character string specifying the three-letter country code of the legislature for which data shall be fetched. Currently one of 'aut', 'can', 'cze', 'esp', 'fra', 'deu', 'irl', 'sco', 'gbr', 'usa\_house', or 'usa\_senate'.

**Format**

Data frame with columns (varies by legislature):

- wikidataid: Wikidata ID identifying a legislator's Wikidata entry (of class 'character').
- parlid: Parliamentary website or website ID (of class 'character').
- btvote: ID for BTVote data on all roll call votes taken in the German Bundestag from 1949 to 2013 and for Reelection Prospects data. The respective ID in BTVote and Reelection Prospects data is 'mp\_id' (of class 'character').
- parlspeech: ID for ParlSpeech datasets containing full-text vectors of plenary speeches. The respective ID in ParlSpeech data is 'speaker' (of class 'character').
- dpsi: ID for Database of Parliamentary Speeches in Ireland. The respective ID in the Database of Parliamentary Speeches in Ireland is 'memberID' (of class 'character').
- eggesspirling: ID for Eggers and Spirling British political development database. The respective ID in Eggers and Spirling data is 'member.id' (of class 'character').
- bioguide: ID for the Voteview congressional roll-call votes database. The respective ID in Voteview data is 'bioguide\_id' (of class 'character').
- icpsr: ID for Congressional Bills Project database. The respective ID in Congressional Bills Project data is 'icpsr' (of class 'character').
- sycomore: Sycomore database of French MPs (of class 'character').
- libcon: Library of Congress ID (of class 'character').
- gnd: German National Library ID (of class 'character').
- bnf: French National Library ID (of class 'character').
- freebase: Freebase ID (of class 'character').
- munzinger: Munzinger archive ID (of class 'character').
- nndb: Notable Names Database ID (of class 'character').
- imdb: Internet Movie Database ID (of class 'character').
- britannica: Encyclopedia Britannica ID (of class 'character').
- quora: Quora ID (of class 'character').
- votesmart: Project Votesmart ID (of class 'character').
- fec: Federal Election Commission ID (of class 'character').
- ballotpedia: Ballotpedia ID (of class 'character').
- opensecrets: Opensecrets ID (of class 'character').
- genealogists: Genealogists ID (of class 'character').
- politfacts: Politfacts ID (of class 'character').
- nkcr: Czech National Library ID (of class 'character').
- parlbio: parliament.uk biography ID (of class 'character').
- parlthesaurus: UK Parliament thesaurus ID (of class 'character').
- national: UK National Archives ID (of class 'character').
- hansard: Hansard (1803-2005) ID (of class 'character').
- publicwhip: PublicWhip ID (of class 'character').
- theyworkforyou: TheyWorkForYou ID (of class 'character').

**Value**

A data frame with columns as specified above.

**Source**

Wikidata API, <https://www.wikidata.org/>

**Examples**

```
# Get entire 'IDs' table for the for the United States Senate
usa_ids <- get_ids(legislature = "usa_senate")
tibble::glimpse(usa_ids)

# Get ICPSR IDs and add 'Offices' table for the United States House
usa_ids_subset <- dplyr::inner_join(x = dplyr::filter(usa_ids,
                                                    !is.na(icpsr)),
                                   y = get_office(legislature = "usa_senate"),
                                   by = "wikidataid")

tibble::glimpse(usa_ids_subset)
```

---

get\_office

*Fetch 'Offices' table*

---

**Description**

Fetches political and other offices of legislators for the specified legislature. Requires a working Internet connection.

**Usage**

```
get_office(legislature)
```

**Arguments**

**legislature** A character string specifying the three-letter country code of the legislature for which data shall be fetched. Currently one of 'aut', 'can', 'cze', 'esp', 'fra', 'deu', 'irl', 'sco', 'gbr', 'usa\_house', or 'usa\_senate'.

**Format**

Data frame in wide format with columns (varies by legislature):

- wikidataid: Wikidata ID identifying a legislator's Wikidata entry (of class 'character').
- office\_1: political or other office held by a legislator (of class 'logical').
- office\_2: ... (of class 'logical').
- ...

**Value**

A data frame with columns as specified above.

**Source**

Wikidata API, <https://www.wikidata.org/>

**Examples**

```
# Get entire 'Offices' table for the United States Senate
usa_offices <- get_office(legislature = "usa_senate")
tibble::glimpse(usa_offices)

# Get 'Offices' table for male members of the United States Senate
usa_offices_subset <- dplyr::semi_join(x = usa_offices,
                                       y = dplyr::filter(get_core(legislature = "usa_senate"),
                                                         sex == "female"),
                                       by = "wikidataid")

tibble::glimpse(usa_offices_subset)
```

---

get_political	<i>Fetch 'Political' table</i>
---------------	--------------------------------

---

**Description**

Fetches political data of legislators for the specified legislature. Requires a working Internet connection.

**Usage**

```
get_political(legislature)
```

**Arguments**

**legislature** A character string specifying the three-letter country code of the legislature for which data shall be fetched. Currently one of 'aut', 'can', 'cze', 'esp', 'fra', 'deu', 'irl', 'sco', 'gbr', 'usa\_house', or 'usa\_senate'.

**Format**

Data frame in long format with columns (varies by legislature):

- **pageid**: Wikipedia page ID identifying a legislator's Wikipedia biography (of class 'integer').
- **session**: Legislative period (of class 'integer').
- **party**: A legislator's party affiliation (of class 'character'). See <https://github.com/saschagobel/legislatoR> for the full form of abbreviated party names and english translations of non-english party names



- constituency: A legislator's constituency (of class 'character').
- constituency2: A legislator's constituency (upper level, if applicable, of class 'character').
- constituency\_id: ID of a legislator's constituency (of class 'character').
- session\_start: Date the legislative period started (of class 'Date').
- session\_end: Date the legislative period ended (of class 'Date').
- service: A legislator's period of service in days during the respective session (of class 'integer').
- government (or similar): Indicator of a legislator's majority status in parliament (of class 'logical'). Further columns with extensions of this might exist.
- leader (or similar): Indicator of a legislator's leader status in parliament (of class 'logical'). Further columns with extensions of this might exist.

### Value

A data frame with columns as specified above.

### Source

Wikipedia, <https://wikipedia.org/>

Czech Republic Parliamentary Members Archive <http://public.psp.cz/sqw/fsnem.sqw?zvo=1>

Spain Parliamentary Members Archive <http://www.congreso.es/portal/page/portal/Congreso/Congreso/Diputados>

### Examples

```
# Get entire 'Political' table for the Czech Poslanecka Snemovna
cze_political <- get_political(legislature = "cze")
tibble::glimpse(cze_political)

# Get 'Political' table for female DSP party members of the Czech Poslanecka Snemovna
cze_political_subset <- dplyr::semi_join(x = dplyr::filter(cze_political,
                                                         party == "ODS"),
                                       y = dplyr::filter(get_core(legislature = "cze"),
                                                         sex == "female"),
                                       by = "pageid")
tibble::glimpse(cze_political_subset)
```

---

get\_portrait

*Fetch 'Portrait' table*

---

### Description

Fetches portrait urls of legislators for the specified legislature. Requires a working Internet connection.

**Usage**

```
get_portrait(legislature)
```

**Arguments**

`legislature` A character string specifying the three-letter country code of the legislature for which data shall be fetched. Currently one of 'aut', 'can', 'cze', 'esp', 'fra', 'deu', 'irl', 'sco', 'gbr', 'usa\_house', or 'usa\_senate'.

**Format**

Data frame with columns (varies by legislature):

- `pageid`: Wikipedia page ID identifying a legislator's Wikipedia biography (of class 'integer').
- `image_url`: URL linking to a legislator's portrait on Wikimedia Commons (of class 'character').

**Value**

A data frame with columns as specified above.

**Source**

Wikipedia API, <https://wikipedia.org/w/api.php>  
 Wikimedia Commons, <https://commons.wikimedia.org/>

**Examples**

```
# Get entire 'Portraits' table for the United States Senate
usa_portraits <- get_portrait(legislature = "usa_senate")
tibble::glimpse(usa_portraits)

# Get 'Portraits' table for Democratic members of the United States Senate
usa_port_subset <- dplyr::semi_join(x = usa_portraits,
                                   y = dplyr::filter(get_political(legislature = "usa_senate"),
                                                       party == "D"),
                                   by = "pageid")
tibble::glimpse(usa_port_subset)
```

---

`get_profession`      *Fetch 'Professions' table*

---

**Description**

Fetches occupational data of legislators for the specified legislature. Requires a working Internet connection.

**Usage**

```
get_profession(legislature)
```

**Arguments**

`legislature` A character string specifying the three-letter country code of the legislature for which data shall be fetched. Currently one of 'aut', 'can', 'cze', 'esp', 'fra', 'deu', 'irl', 'sco', 'gbr', 'usa\_house', or 'usa\_senate'.

**Format**

Data frame in wide format with columns (varies by legislature):

- `wikidataid`: Wikidata ID identifying a legislator's Wikidata entry (of class 'character').
- `occupation_1`: occupation a legislator practiced or was trained in (of class 'logical').
- `occupation_2`: ... (of class 'logical').
- ...

**Value**

A data frame with columns as specified above.

**Source**

Wikidata API, <https://www.wikidata.org/>

**Examples**

```
# Get entire 'Professions' table for the United States House
usa_offices <- get_office(legislature = "usa_house")
tibble::glimpse(usa_offices)

# Get 'Professions' table for female members of the United States House
usa_offices_subset <- dplyr::semi_join(x = usa_offices,
                                       y = dplyr::filter(get_core(legislature = "usa_house"),
                                                         sex == "female"),
                                       by = "wikidataid")
tibble::glimpse(usa_offices_subset)
```

---

get_social	<i>Fetch 'Social' table</i>
------------	-----------------------------

---

### Description

Fetches social media handles and website URLs of legislators for the specified legislature. Requires a working Internet connection.

### Usage

```
get_social(legislature)
```

### Arguments

`legislature` A character string specifying the three-letter country code of the legislature for which data shall be fetched. Currently one of 'aut', 'can', 'cze', 'esp', 'fra', 'deu', 'irl', 'sco', 'gbr', 'usa\_house', or 'usa\_senate'.

### Format

Data frame with columns (varies by legislature):

- `wikidataid`: Wikidata ID identifying a legislator's Wikidata entry (of class 'character').
- `twitter`: Twitter handle (of class 'character').
- `facebook`: Facebook handle (of class 'character').
- `youtube`: Youtube ID (of class 'character').
- `googlep`: Google Plus ID (of class 'character').
- `instagram`: Instagram handle (of class 'character').
- `linkedin`: LinkedIn ID (of class 'character').
- `website`: Personal website URL (of class 'character').

### Value

A data frame with columns as specified above.

### Source

Wikidata API, <https://www.wikidata.org/>

## Examples

```
# Get entire 'Social' table for the UK House of Commons
gbr_social <- get_social(legislature = "gbr")
tibble::glimpse(gbr_social)

# Get 'Social' table for members of the UK House of Commons with available TheyWorkForYou ID
gbr_social_subset <- dplyr::semi_join(x = gbr_social,
                                     y = dplyr::filter(get_ids(legislature = "gbr"),
                                                         !is.na(theyworkforyou)),
                                     by = "wikidataid")
tibble::glimpse(gbr_social_subset)
```

---

get_traffic	<i>Fetch Wikipedia 'Traffic' table</i>
-------------	--

---

## Description

Fetches daily user traffic on legislators' Wikipedia biographies for the specified legislature. Requires a working Internet connection.

## Usage

```
get_traffic(legislature)
```

## Arguments

**legislature** A character string specifying the three-letter country code of the legislature for which data shall be fetched. Currently one of 'aut', 'can', 'cze', 'esp', 'fra', 'deu', 'irl', 'sco', 'gbr', 'usa\_house', or 'usa\_senate'.

## Format

Data frame with columns:

- **pageid**: Wikipedia page ID identifying a legislator's Wikipedia biography (of class 'integer').
- **date**: Date for which user traffic is recorded, from 2015-07-01 to 2018-12-31 UTC (of class 'POSIXct').
- **traffic**: Daily non-unique user visits (of class 'numeric').

## Value

A data frame with columns as specified above.

## Source

Wikimedia API, [https://wikimedia.org/api/rest\\_v1/http://petermeissner.de:8880/](https://wikimedia.org/api/rest_v1/http://petermeissner.de:8880/)

## Examples

```
# Get entire 'Traffic' table for the Scottish Parliament
sco_traffic <- get_traffic(legislature = "sco")
tibble::glimpse(sco_traffic)

# Add Wikidataid to 'Traffic' table for the Scottish Parliament
sco_traffic_subset <- dplyr::inner_join(x = dplyr::select(get_core(legislature = "sco"),
                                                         pageid, wikidataid),
                                       y = sco_traffic,
                                       by = "pageid")

tibble::glimpse(sco_traffic_subset)
```

---

legislatoR

*legislatoR*

---

## Description

Facilitates access to the Comparative Legislators Database (CLD). The CLD includes political, sociodemographic, career, online presence, public attention, and visual information for over 45,000 contemporary and historical politicians from ten countries.

## legislatoR functions

[cld\\_content](#): returns a named list of legislatures and sessions available in the CLD.

[get\\_core](#): fetches sociodemographic data of legislators.

[get\\_political](#): fetches political data of legislators.

[get\\_history](#): fetches full revision histories of legislators' Wikipedia biographies.

[get\\_traffic](#): fetches daily user traffic on legislators' Wikipedia biographies.

[get\\_social](#): fetches social media handles and website URLs of legislators.

[get\\_portrait](#): fetches portrait urls of legislators.

[get\\_office](#): fetches political and other offices of legislators.

[get\\_profession](#) fetches occupational data of legislators.

[get\\_ids](#): fetches a range of IDs of legislators.

# Index

`cld_content`, [2](#), [14](#)

`get_core`, [3](#), [14](#)

`get_history`, [4](#), [14](#)

`get_ids`, [5](#), [14](#)

`get_office`, [7](#), [14](#)

`get_political`, [8](#), [14](#)

`get_portrait`, [9](#), [14](#)

`get_profession`, [10](#), [14](#)

`get_social`, [12](#), [14](#)

`get_traffic`, [13](#), [14](#)

`legislatoR`, [14](#)

`legislatoR-package (legislatoR)`, [14](#)