# PolDem – Protest Dataset

# 30 European countries

## - Version 1 -

poldem-protest_30

***Please cite as:***

Kriesi, Hanspeter, Wüest, Bruno, Lorenzini, Jasmine, Makarov, Peter, Enggist, Matthias, Rothenhäusler, Klaus, Kurer, Thomas, Häusermann, Silja, Patrice Wangen, Altiparmakis, Argyrios, Borbáth, Endre, Bremer, Björn, Gessler, Theresa, Hunger, Sophia, Hutter, Swen, Schulte-Cloos, Julia, and Wang, Chendi. 2020. *PolDem-Protest Dataset 30 European Countries*, Version 1.

Short version for in-text citations & references: Kriesi, Hanspeter et al. 2020. *PolDem-Protest Dataset 30 European Countries*, Version 1.

[www.poldem.eu](http://www.poldem.eu)

**Overview**

This document contains

(1) information on the variables included in the protest event data compiled by the POLCON and Years-of-Turmoil research teams. The data collection relies on a semi-automated coding of newswire reports and covers 30 European countries and the time period from 2000 until 2015. The dataset is currently updated.

(2) A detailed documentation of the data collection process (p. 12ff.)

## Commented list of indicators

*Files:* PEA30sixteen_Events_12042018.dta and PEA30sixteen_Events_12042018.csv
*N observations:* 30,852
*N variables:* 53

| Variable name | Values | Description |
|---|---|---|
| id_doc | 1…18222 | document identifier (unambiguously describes a single document) |
| id_coder | 1…34 | coder identifier (anonymized) |
| id_event | 1…30852 | event identifier (unambiguously describes one protest event, i.e. a unique combination of date, action and location) |
| entry_time | "2016-05-03 19:59:46" … | date and time of event annotation |
| doc_country | 1 = "austria"<br>…<br>30 = "united kingdom" | country assigned to documents during preprocessing |
| doc_publdate | "2000-09-12" … | publication date of document |
| doc_source | 1 = "AFP"<br>…<br>11 = "PAP" | newswire that published the document |
| doc_probability | 0.6100178…1 | probability that document includes protest events, estimated during preprocessing |
| event_index | 1…23 | index of events annotated per document |
| event_n | 1…4 | number of observations of an event (before duplicate removal) |
| country_name | 1 = "austria"<br>…<br>31 = "northern Ireland" | country of event (Northern Ireland is listed separately) |
| country_size | 0.3…80.1 | size of population, in million inhabitants |
| community | "vienna" … | community in which event took place |
| date | "2000-01-01" … | date of protest event (starting date if time period was coded) |
| month | 1…12 | month in which event happened |

| | | |
|---|---|---|
| year | 2000…2015 | year in which event happened |
| period | 0 = "normal times"<br>…<br>3 = "refugee crisis" | time periods related to the Great Recession |
| action_form | 1 = "petitions, symbolic actions"<br>…<br>6 = "other protest" | type of action used at protest event |
| radical_action | 0 = "not radical"<br>1 = "confrontational"<br>2 = "violent" | radical action forms (confrontations and violence) used at protest event |
| part_all | 0…1000000 | recoded number of participants with estimates for missing values |
| big_event | 0 = "no"<br>1 = "yes" | events with exceptionally strong mobilization |
| issue_econ_private | 0 = "no"<br>1 = "yes" | economic claims addressed to firms/employers |
| issue_econ_public | 0 = "no"<br>1 = "yes" | economic claims addressed to public institutions |
| issue_cult_lib | 0 = "no"<br>1 = "yes" | cultural liberalist issues |
| issue_environment | 0 = "no"<br>1 = "yes" | environmental issues |
| issue_political | 0 = "no"<br>1 = "yes" | political issues |
| issue_regional | 0 = "no"<br>1 = "yes" | regionalist issues |
| issue_cult_cons | 0 = "no"<br>1 = "yes" | cultural conservative issues |
| issue_xeno | 0 = "no"<br>1 = "yes" | xenophobia |
| issue_other | 0 = "no"<br>1 = "yes" | other issues |
| issue_dontknow | 0 = "no"<br>1 = "yes" | an issue is specified but attribution is unclear |
| issue_missing | 0 = "no"<br>1 = "yes" | issue for protest event is missing |
| actor_party_left | 0 = "no"<br>1 = "yes" | left parties |

| | | |
|---|---|---|
| actor_party_right | 0 = "no"<br>1 = "yes" | right parties |
| actor_party_unknown | 0 = "no"<br>1 = "yes" | parties with unclear ideological leaning |
| actor_union_private | 0 = "no"<br>1 = "yes" | private sector union |
| actor_union_public | 0 = "no"<br>1 = "yes" | public sector union |
| actor_union_both | 0 = "no"<br>1 = "yes" | union representing the public and private sector |
| actor_union_unknown | 0 = "no"<br>1 = "yes" | union with unknown sector affiliation |
| actor_group_pens | 0 = "no"<br>1 = "yes" | social group representing pensioners |
| actor_group_stud | 0 = "no"<br>1 = "yes" | social group representing students |
| actor_group_occup | 0 = "no"<br>1 = "yes" | social group representing occupations |
| actor_group_other | 0 = "no"<br>1 = "yes" | social group representing others |
| actor_other | 0 = "no"<br>1 = "yes" | other organized actors |
| actor_dontknow | 0 = "no"<br>1 = "yes" | an actor is specified but attribution is unclear |
| actor_missing | 0 = "no"<br>1 = "yes" | no actors recorded for a protest event |
| weight_sample | 1 = ".25 sample"<br>2 = ".5 sample"<br>3 = "1.0 sample" | weight controlling for different document sampling probabilities |
| weight_newswire | 0.5…1 | weight controlling for newswire focus on single countries |
| weight_country | 1.15…5.383276 | weight controlling for different country sizes (1 / logged population size) |
| weighted_event | 0.119537…3.383174 | event occurrence with sample, newswire and country weights applied |
| weighted_part_all | 0…608695.7 | number of participants with sample, newswire and country weights applied |

**Descriptive statistics on some of the main indicators**

The ranking of countries (indicator: 'country_name') in terms of the numbers of protest events is shown in the following table.

| **Country** | **Freq.** | **Percent** |
|---|---:|---:|
| *United Kingdom* | 3,826 | 12.40 |
| *Greece* | 3,653 | 11.84 |
| *France* | 3,147 | 10.2 |
| *Spain* | 2,520 | 8.17 |
| *Italy* | 2,299 | 7.45 |
| *Germany* | 2,252 | 7.30 |
| *Czech Republic* | 2,164 | 7.01 |
| *Austria* | 905 | 2.93 |
| *Poland* | 842 | 2.73 |
| *Belgium* | 766 | 2.48 |
| *Romania* | 744 | 2.41 |
| *Lithuania* | 665 | 2.16 |
| *Bulgaria* | 640 | 2.07 |
| *Sweden* | 634 | 2.05 |
| *Portugal* | 598 | 1.94 |
| *Slovakia* | 594 | 1.93 |
| *Latvia* | 582 | 1.89 |
| *Switzerland* | 553 | 1.79 |
| *Hungary* | 526 | 1.70 |
| *Netherlands* | 502 | 1.63 |
| *Estonia* | 408 | 1.32 |
| *Denmark* | 381 | 1.23 |
| *Cyprus* | 370 | 1.20 |
| *Finland* | 358 | 1.16 |
| *Ireland* | 313 | 1.01 |
| *Slovenia* | 263 | 0.85 |
| *Norway* | 215 | 0.70 |
| *Iceland* | 67 | 0.22 |
| *Malta* | 35 | 0.11 |
| *Luxembourg* | 30 | 0.10 |
| *Total* | 30,852 | 100 |

The most important observation is that some large countries, most notably the United Kingdom, come out as top ranked countries not because the protest activity in these countries is exceptionally high, but because they are overrepresented in our sample of newswires. For the UK the conflict in Northern Ireland additionally increases the level of protests. This led us a) to conduct many analyses with Northern Ireland as a separate geographical entity (see indicator 'country_name'), and b) to apply country specific weights (see the indicator 'weight_country' as well as the sections on weighting in the documentation, PEA30sixteen_Documentation_12-4-2018.docx).

Among the top locations (the top 20 are shown in the following table; indicator: 'community'), there are almost exclusively capital cities – the exceptions being Belfast and Thessaloniki. Athens by far is the city with most protests, since protest in the Greek capital make up for about 6 per cent of all events. Hence, our data is skewed towards the major European cities, which should not come as a surprise given that we are relying on English-language newswires publishing mainly for an international audience. In the documentation (PEA30sixteen_Documentation_12-4-2018.docx), we report tests on other potential biases related to specific protest event indicators (i.e. action forms, issues and actors). This bias towards the capitals, however, is the only significant one.

| **community** | **Freq.** | **Percent** |
|---|---|---|
| *Athens* | 1,792 | 5.81 |
| *London* | 1,156 | 3.75 |
| *Prague* | 1,146 | 3.71 |
| *Paris* | 905 | 2.93 |
| *Rome* | 726 | 2.35 |
| *Madrid* | 687 | 2.23 |
| *Belfast* | 576 | 1.87 |
| *Brussels* | 479 | 1.55 |
| *Berlin* | 466 | 1.51 |
| *Riga* | 422 | 1.37 |
| *Thessaloniki* | 385 | 1.25 |
| *Vilnius* | 383 | 1.24 |

| | | |
|---|---|---|
| *Vienna* | 378 | 1.23 |
| *Sofia* | 358 | 1.16 |
| *Warsaw* | 337 | 1.09 |
| *Budapest* | 304 | 0.99 |
| *Bucharest* | 291 | 0.94 |
| *Stockholm* | 268 | 0.87 |
| *Bratislava* | 251 | 0.81 |
| *Tallinn* | 251 | 0.81 |

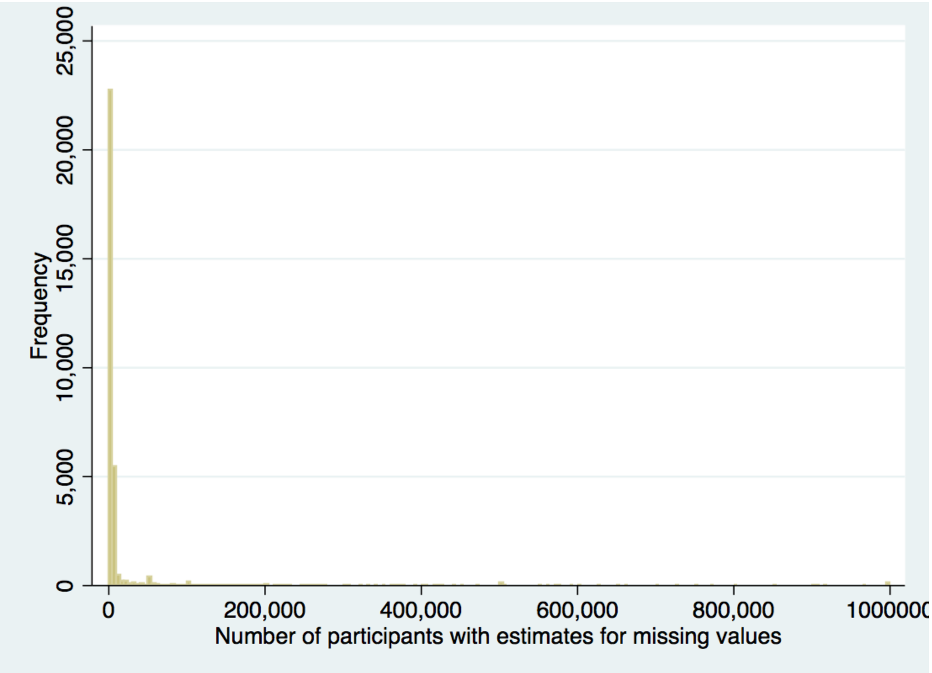
If we break down the events by year (indicator: 'year'; see next table), we see that the range across the years is rather large (minimum of 1,297; maximum of 2,586) and that we have three peaks. The single years of 2000 (global justice movement) and 2005 (French riots and Iraq war protests) as well as the extended period of six years from 2007 until 2012 most likely linked to the Great Recession and the Euro crisis.

| **Year** | **Freq.** | **Percent** |
|---|---|---|
| *2000* | 2,586 | 8.38 |
| *2001* | 1,770 | 5.74 |
| *2002* | 1,756 | 5.69 |
| *2003* | 1,926 | 6.24 |
| *2004* | 1,478 | 4.79 |
| *2005* | 2,012 | 6.52 |
| *2006* | 1,792 | 5.81 |
| *2007* | 2,031 | 6.58 |
| *2008* | 2,473 | 8.02 |
| *2009* | 2,189 | 7.10 |
| *2010* | 2,167 | 7.02 |
| *2011* | 2,104 | 6.82 |
| *2012* | 2,036 | 6.60 |
| *2013* | 1,732 | 5.61 |
| *2014* | 1,297 | 4.20 |
| *2015* | 1,503 | 4.87 |
| *Total* | 30,852 | 100 |

Peaceful demonstrations clearly are most prevalent among the different action forms (see table below; indicator: 'action_form'). Violent protests and strikes are rather frequent as well, but nevertheless, their share is less than half than the share of demonstrations.

| Action form | Freq. | Percent |
|---|---|---|
| *demonstrations* | 13,327 | 43.2 |
| *violent protest* | 6,224 | 20.17 |
| *strikes* | 5,177 | 16.78 |
| *confrontations, blockades* | 2,594 | 8.41 |
| *petitions, symbolic actions* | 2,450 | 7.94 |
| *other protest* | 1,080 | 3.5 |
| *Total* | 30,852 | 100 |

As for the number of participants (indicator: 'part_all'), the following figure shows that almost all events have either very few participants or no participation is recorded. There are, however, 787 big events (indicator: 'big_events'). These are events, which had over 100'00 participants for countries with a population of more than 5 million inhabitants, or which had a participation rate that was above the threshold of the median plus one standard deviation for small countries with less than 5 million inhabitants.




Economic issues, and especially economic claims addressed to public institutions, are prevalent in our corpus (indicators: 'issue_econ_private', 'issue_econ_public', 'issue_cult_lib', 'issue_environment', 'issue_political', 'issue_regional', 'issue_cult_cons', 'issue_xeno', 'issue_other', 'issue_dontknow', and ‘issue_missing’). This is of course lined to the Great Recession and the following Euro crisis, which certainly have increased the salience of these issues. Also, only for about four percent of the events, no issue was annotated.

| Issue | Freq. | Percent[1] |
|---|---|---|
| *public economic* | 9,247 | 29.97 |
| *cultural liberal* | 5,240 | 16.98 |
| *political* | 4,678 | 15.16 |
| *private economic* | 3,769 | 12.22 |
| *regional* | 2,212 | 7.17 |
| *environment* | 2,051 | 6.65 |
| *xenophobia* | 1,806 | 5.85 |
| *cultural conservative* | 903 | 2.93 |
| *other issue* | 3,339 | 10.82 |
| *no issue* | 1,304 | 4.23 |
| *Total* | 30,852 | |

[1] The shares do not sum up to 100 percent since more than one issues could be coded for each event.

Lastly, as far as the actors are concerned (indicators: ‘actor_party_left’, ‘actor_party_right’, ‘actor_party_unknown’, ‘actor_union_private’, ‘actor_union_public’, ‘actor_union_both’, ‘actor_union_unknown’, ‘actor_group_pens’, ‘actor_group_stud’, ‘actor_group_occup’, ‘actor_group_other’, ‘actor_other ‘actor_dontknow’ and ‘actor_missing’), we have far more missing information. For about 35 percent of all events, no information on organized interests was found. Most notably the involvement of political parties in contentious mobilization is below expectations.

| Actors | | Freq. | Percent[1] |
|---|---|---|---|
| parties | left | 941 | 3.05 |
| | right | 940 | 3.05 |
| | other | 236 | 0.76 |
| unions | private sector | 454 | 1.47 |
| | public sector | 993 | 3.22 |
| | private and public | 2,078 | 6.74 |
| | others | 2,229 | 7.22 |
| Social groups | pensioners | 103 | 0.33 |

| | | | |
|---|---|---|---|
| | students | 1,180 | 3.82 |
| | occupations | 5,236 | 16.97 |
| | other groups | 2,402 | 7.79 |
| Other actors | | 5,676 | 18.4 |
| No actor | | 10,867 | 35.22 |
| Total | | 30,852 | |

[1] The shares do not sum up to 100 percent since more than one issues could be coded for each event.

***Protest Event Data on 30 European Countries 2000-2015***

# Documentation

16.4.2018

---

## Contents



---

## Abstract


This document entails information on the production of the Protest Event Analysis (PEA) of newswire reports that cover 30 European countries and the time period from 2000 until 2015. All work was done by the POLCON and Years-of-Turmoil research teams from September 2014 until April 2018.


If you can make use of the data, please cite it as follows:

Kriesi, Hanspeter, Wüest, Bruno, Lorenzini, Jasmine, Makarov, Peter, Enggist, Matthias, Rothenhäusler, Klaus, Kurer, Thomas, Häusermann, Silja, Patrice Wangen, Altiparmakis, Argyrios, Borbáth, Endre, Bremer, Björn, Gessler, Theresa, Hunger, Sophia, Hutter, Swen, Schulte-Cloos, Julia, and Wang, Chendi. 2020. *PolDem-Protest Dataset 30 European Countries*, Version 1.

## 2. Introduction

The analysis of contentious politics has a long tradition in the social sciences. One of the key methods in this field is protest event analysis, a form of content analysis that allows to systematically collect, quantify, and process large amounts of information on political protest across time and countries (Hutter 2014, 2014a). Protest event data has usually relied on newspaper coverage as data source[1], and, mostly on manual coding. Yet, the manual coding of protest events across time and space is intensely time consuming. This posed a particular challenge for our research design, because we intended to cover the EU-27 countries plus three non-EU members (Iceland, Norway, and Switzerland) over a 16-years period (2000-2015).

The unit of observation for our study is a protest event, defined by a time, a location, and a particular protest action form. Our final dataset includes approximately 31'000 protest events across all 30 countries and 16 years of observation. For the collection of such a large data set we chose to rely on a combination of automated and manual coding, given that manual coding alone was not feasible. Following the suggestions by Nardulli et al. (2015), we ended up choosing a hybrid procedure that combines the strengths of machine learning and human coding. More precisely, we develop a semi-automatic procedure, in which an extensive application of Natural Language Processing (NLP) tools pre-selects the documents for the subsequent human coding of the fine-grained protest event data. The NLP tool pipeline allows us to broaden the data basis to millions of news reports, and to extend the sample of countries, years, and sources beyond what is usually possible in purely manual approaches. However, it also required the linguistic homogeneity of our sources. Following the lead of Beissinger and Sasse (2014), we chose to retrieve and code protest events reported

[1] Sometimes, police records, social movement documents, or other written material documenting contentious political actions are used as well (McCarthy et al. 1996, Oliver and Mahoney 2000, Wouters 2013).

by 10 English-speaking news agencies – all we could find in the 30 countries covered by our study.

The aim of this documentation is to highlight the technical aspects of the data generation process in order to facilitate the transfer of tools and methods to other researchers engaged in the semi-automated coding of protest events. This document is structured as follows. The procedure for the automated selection of news sources on protest events is described in section 2, while the manual coding of protest events is described in section 3. In section 4, we present the data by providing descriptive analyses of selected indicators. Section 5, finally, entails a discussion of how the protest event data can be weighted in order to obtain a data set that can be used to compare contentious mobilization across the countries and years of the analysis.

---

## 2. Automated selection of protest documents

The starting point of the analysis are ten English-language newswire agencies[2] that are accessible at the Lexis-Nexis data service. We aim at including as many news agencies as possible in order to cover as comprehensible as possible the different regions and countries included in our study (Jenkins and Maher 2016). With a search query that comprises about forty keywords[3], we retrieve an initial set of 5.2 million news reports published in the sixteen-year period from 2000 to 2015. We deliberatively opt for a query that retrieves as many relevant documents as possible in order to minimize the risk of false negatives, even at

[2] We include the following news agencies: AFP, AP, APA, BBC, BNS, CTK, DPA, MTI, PA, and PAP. Our goal was to include not only the major news agencies (AFP, DPA, PA) but also some regional ones covering Eastern and Southern Europe more in depth.

[3] Query string "initiative OR referendum OR petition! OR signature! OR campaign! OR protest! OR demonstrat! OR manifest! OR marche! OR marchi! OR parade OR rall! OR picket! OR (human chain) OR riot! OR affray OR festival OR ceremony OR (street theatre) OR (road show) OR vigil OR strike! OR boycott! OR block! OR sit-in OR squat! OR mutin! OR bomb! OR firebomb! OR molotov OR graffiti OR assault OR attack OR arson OR incendiar! OR (fire I/1 raising) OR (set AND ablaze) OR landmine OR sabot! OR hostage! OR assassinat! OR shot OR murdered OR killed"

the expense of obtaining a large share of irrelevant documents that need to be filtered out in later stages.

The drawback is that our search is ‘greedy’, with the consequence that it results in only about 5 percent of documents actually reporting about protest events in the countries of interest. The low share of true positives accordingly is one of the biggest challenges for both the automated as well as the manual coding of protest events. To go from the 5.2 million reports to a set that can be coded manually without losing too much relevant reports, we implemented the following steps: 1) removal of duplicated reports, 2) discard reports not related to our countries of interest using a meta-data filter, 3) apply a supervised document classifier distinguishing relevant reports from irrelevant ones, 4) discard textually very similar reports, and 5) apply a supervised protest mention classifier.

Table 1 summarizes the steps of this automated pipeline to filter irrelevant documents and indicates our progress in terms of reducing the number of reports to an amount that can be handled by human coders. In the following, we not only provide a description as well as a validation of every step.

Table 1: *Steps of the automated filtering of irrelevant documents*

| | **N reports in corpus** | **Reduction in %** |
|---|---|---|
| *Keyword search* | 5,251,894 | |
| *Duplicate removal* | 4,211,759 | 19.8 |
| *Location-based filter* | 1,116,337 | 73.5 |
| *Document classifier* | 157,572 | 85.9 |
| *Near-duplicate removal* | 147,846 | 6.2 |
| *Protest mention classifier* | 101,877 | 31.1 |
| *Country-specific sampling* | 45,680 | 44.8 |

*Duplicate removal*

Since the different keywords in our search query sometimes match the same documents, we had to remove about 20 percent of the originally retrieved documents since they were duplicated. This could easily be achieved by relying on the unique document identifiers provided by Lexis-Nexis.

*Location-based filtering*

Besides duplicates, certain news reports can also be irrelevant in terms of the geographic scope of our analysis. The determination whether a story is relevant for any of our thirty countries is straightforward, since virtually all news reports are labelled with metadata that also includes a list of countries associated with the contents of the report. Hence, we can filter out documents that do not feature any of our countries of interest. For countries for which we already have a very high retrieval rate (e.g. the UK, France, and Germany), we additionally discard documents in which this country is not among the top-ranked and in which no other relevant country was found. Thus, we aim at up-weighting the share of sampled documents for countries where only few news reports are available, which are Austria, Belgium, Bulgaria, Switzerland, Cyprus, Denmark, Finland, Greece, Island, Luxembourg, Malta, Netherlands, Portugal, Romania, Slovakia, Slovenia and Sweden.

The location-based filtering just described reduces the corpus considerably, with about three quarters of documents removed (73.5 percent). The lion's share of this reduction (about 70 of these 73.5 percent) is due to the removal of documents that do not feature any of our 30 countries of interest. Thus, only 3.5 percent of the documents are discarded because they only cover one or more large countries. Nevertheless, the share of documents for countries with a high original retrieval rate, namely Germany, the UK, France, Spain, Italy, and Ireland decreases a little[4].

The up-weighting of smaller countries still has the intended effect, since it slightly increases their share in all documents. Moreover, we can show that this up-weighting does not induce a severe bias into the corpus. The Kullback-Leibler distance[5] between the distribution of the number of documents across the 30 countries before and after the application of the location-based filtering is 0.006 for all countries, 0.004 for the big countries and even 0.000001 for the small countries[6]. Other subsequent steps of our semi-automated data generation process such as the classification, document sampling, and manual annotation induce much more divergence in terms of the Kullback-Leibler distance on the distribution of documents across countries.

*Document classification*

For the next step, we train and apply a statistical classifier[7] to the corpus. To this end, we first annotate manually about 7,500 documents into relevant (i.e. mentioning recent protest events in Europe) and irrelevant ones. Second, we model the reports' contents with a classic bag-of-words model. Under this model, a document is represented with a set of features[8] – one for each word or short phrase – whose values are the relative frequencies of these words and phrases in that document (Manning et al. 2008). We additionally apply a number of preprocessing methods that have been empirically shown to produce the most informative bag-of-words representations (Sebastiani 2002). Since the contents of a document depend

[4] In this context, Ireland is not a small country. On the contrary, some of the news agencies we compiled reports from are heavily over-reporting events from the UK and Ireland.
[5] The Kullback-Leibler distance, also called the relative entropy measure, indicates the non-symmetric difference between two distributions. It is always non-negative, whereas values close to zero indicate almost identical distributions.
[6] We measured these distances only for the documents retrieved for the years 2005 to 2014.
[7] In machine learning, a classification model (e.g. logistic regression) is called a classifier. Training a classifier means fitting a classification model to a priori compiled training data.
[8] Independent variables are called features in machine learning (Grimmer and Stewart 2013)

much more on its notional words (nouns, verbs, adjectives) than the functional ones (e.g. articles and prepositions) and punctuation, it is a good idea to remove all functional words and punctuation from the reports before deriving the feature vectors for the training of the classifier. In addition, since most of the words in a collection of documents occur very infrequently, one commonly applied preprocessing technique is to collapse multiple related words to one (e.g. “protested”, “protesters”, “protest”, “protests”, “protesting”). We use stemming to this purpose, which reduces a word to its root by removing endings e.g. the plural and gerund endings “-s” and “-ing”.

Additional to the bag-of-words representation, we introduce features signaling the number of times a document mentions relevant or irrelevant locations. We experimentally find that this helps the classification model generalize better. To this end, we compile lists of European and common non-European country and city names. Also, we find that building a model only over sentences that contain a keyword from the search query (such as “attack”, “protest”, “demonstrate”, etc.) does not impair model generalization, but greatly speeds up training. We always include the words from the title, byline, and lead sentences into the model.

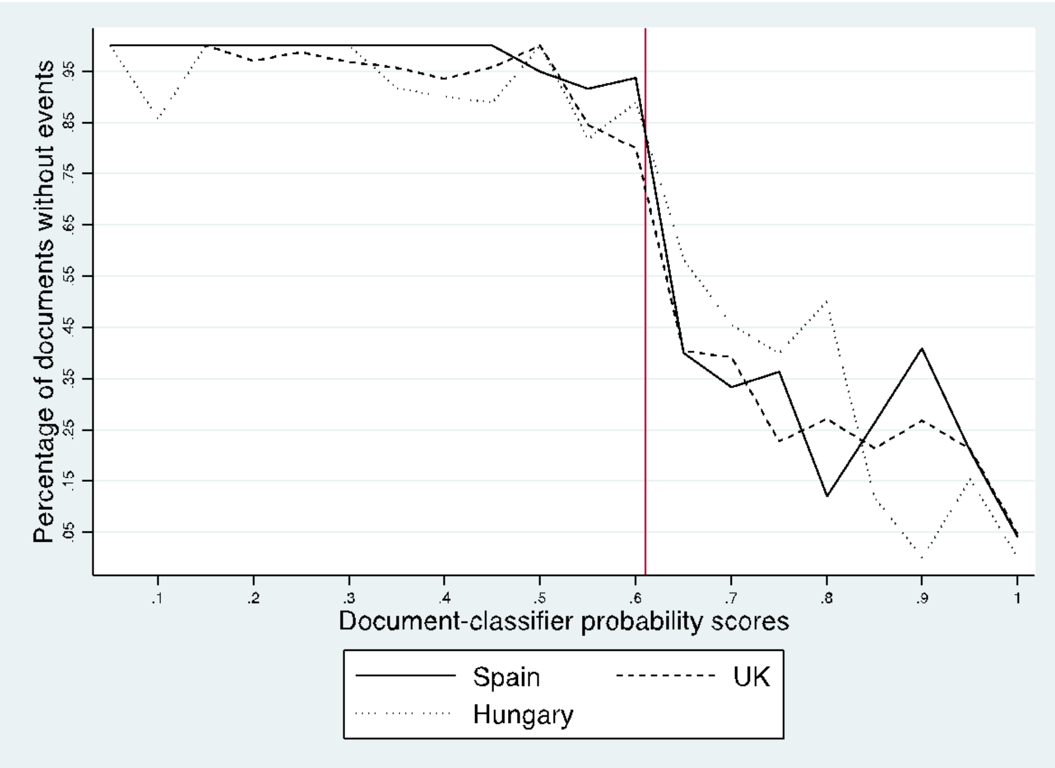
On this set of features, we fit a logistic regression model regularized with the elastic net penalty (Zou & Hastie 2005). The penalty term is there to regularize a model that has many more independent variables than training samples. Furthermore, this penalty has the effect of automatically identifying the most informative independent variables for the inclusion into the classification model and setting the parameters of the remaining variables to zeros. Also, we tune the classification threshold to obtain a higher precision as described in appendix A. We use the implementation of this algorithm from *scikit-learn*, a Python machine learning library (Pedregosa et al. 2011).[9]

One common estimate of how well a classification model generalizes to unseen data (i.e. how well it performs on them in terms of accuracy or some other metric) is its performance on a held-out test set (Friedman et al. 2001). For our document classification model, we report a test-set precision of 0.52 and a recall of 0.66. Precision is the proportion of documents that the model has correctly identified as protest-relevant. Recall indicates the proportion of true protest-relevant documents that the classifier retrieves. The recall of 66 percent is for the whole test set, i.e. for both relevant and irrelevant documents. If recall is calculated for irrelevant documents only, it increases to 94 percent. Hence, this step of our filtering procedure is able to find and discard irrelevant reports in a reliable way, and it leads to the exclusion of about one million of reports.

We are also able to present a more detailed evaluation of whether the document classifier introduces any substantial biases into the corpus of relevant news reports. We do so by comparing the share of irrelevant documents above and below the classification threshold (figure 1) and by comparing the number of events with different action forms below and above the threshold (figure 2).

[9] Our code is available at: https://gitlab.cl.uzh.ch/makarov/polcon/

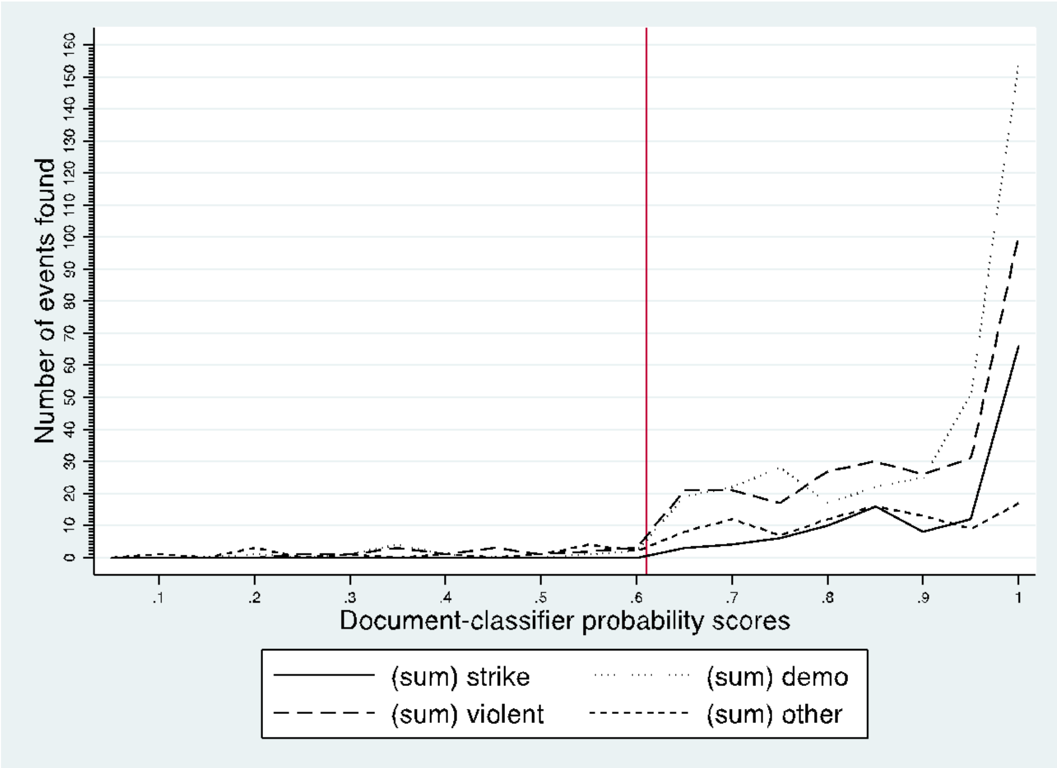
Figure 1: *Share of irrelevant documents above and below the document classifier threshold*




Note: Test includes only data on Hungary, Spain, and UK for the years 2004 and 2012.

In order to perform the two tests, we manually code 1,000 documents that are randomly drawn from all the documents classified as irrelevant for three countries Hungary, Spain, and the UK. These countries belong to the three European regions that we cover. For every country, two years are considered, one pre-crisis (2004) and one post-crisis (2012) year. Figure 1 shows that there are almost no relevant documents below the threshold and that this is the case for all three countries selected for the test. Moreover, above the threshold of 0.61, the share of irrelevant documents rapidly declines for all countries. This is a clear indication that the document classification model is confident of whether documents contain protest events irrespective of the country.

Figure 2: *Number of events found above and below the document classifier threshold by action forms*




Note: Test includes only data on Hungary, Spain, and UK for the years 2004 and 2012.

Using the same set of documents, we assess potential biases in action forms. Figure 2 shows the distribution of events found at the various probability levels for different forms of protest actions–demonstrations, strikes, violent actions, and a residual category including all other forms. Again, very few events are found in documents below the threshold irrespective of the action form. Moreover, the number of events increases rapidly as we move above the threshold for demonstrations, strikes, and violent actions. Across these three forms, most events are found when the document classification model is highly confident that the document reports about protest events (probability scores above 0.90). The trends are similar across the three main action forms. They only differ for the residual category including all other actions. This category contains a large variety of forms and is thus more difficult to identify for the classification model and, at the same time, the least interesting category in our analysis. In general, the results indicate that the document classification model is not biased against any specific form of action.

*Near-duplicate removal*

We also face the problem of very similar reports that only differ in the wording of one or two sentences. A large part of these near-duplicates is due to news agencies re-writing the same news story over the course of the day as events unfold. Furthermore, news agencies also share news reports while modifying only small parts of the reports' content. To remove such near-duplicates, we compare all documents with the SpotSigs algorithm (Theobald et al. 2008). SpotSigs detects near-duplicates in two steps. First, anchor words – in our case, the search terms from the document retrieval – are combined with short chains of adjacent content terms and document meta-data in order to create robust document signatures. Second, documents are matched on the basis of a (multi-)set Jaccard similarity between these document signatures. The Jaccard similarity is defined as the size of the intersection divided by the size of the union of the document signatures.[10] With this approach, we identify groups of news reports with a Jaccard similarity of 0.75 or higher and keep only one report of each group in our corpus. We pick the optimal threshold value by manually evaluating 100 duplicate pairs for each of the following thresholds: 0.65, 0.70, 0.75, and 0.80. For a threshold value of 0.75 or higher, we find no errors. After the document classifier and before the event mention detection, our corpus contains about six percent of such near-duplicate documents, which, accordingly, are discarded.

[10] For our implementation, see https://gitlab.cl.uzh.ch/rothenha/nearDuplicateDetection/

*Event mention detection*

Even after the four filtering steps discussed above, we still have too many reports left, which is why we applied a more fine-grained approach than the document classifier. This approach is based on the observation that the bag-of-words assumption of the document classifier does not adequately address the nuances of our definition of a relevant protest event. If, for instance, the dateline includes "London" and the text contains a phrase "a terrorist attack in Beirut", from the perspective of a bag-of-words model, both place names (London and Beirut) could equally likely name the location of that protest event. That is actually only true in a model where the features are exclusively words. It is not entirely true in our case, since we also use short phrases as predictors. However, in practice, the elastic net penalty drops many of such variables out of the model since the relative frequencies of most phrases correlate poorly with the outcome, i.e. whether the report is relevant or irrelevant.

The solution to such a problem is that a classifier needs to consider the immediate context of words and phrases that likely name protest actions. We therefore fit another classification model for the identification of event mentions in the text. As training data, we use 518 documents annotated at the word level for protest event mentions. Domain experts annotated the documents, i.e. they were asked to mark word spans in the text that they believe most directly denote protest events, e.g. "take to the streets", "violent demonstration", "signed a petition", "carried banners", etc. With this model, we therefore try to predict whether or not some span of words mentions a protest event.

One challenge related to the use of annotated text is the consistency of the data across annotators, which reveals itself in a rather low inter-annotator agreement[11]. In our case, the annotation has been performed in a content-driven fashion with little constraints on what could or could not be annotated. In view of the difficulties, we do not approach event mention detection as a word sequence classification task, despite standard practice (Lafferty et al. 2001). Instead, we simplify the task further into the prediction, for a common noun or verb, of whether or not it syntactically governs a likely protest event phrase (e.g. the noun "riots" in the phrase "violent riots" or "threw" in "threw stones"). In the NLP literature, such a word is called *event trigger* (ACE 2005).

To turn our word span annotations into event trigger annotations, we automatically identify event triggers in annotated word spans using a syntactic dependency analysis tool, the Stanford dependency parser (Manning et al. 2014). As the tool occasionally produces erroneous analyses, we manually went through the resulting event trigger annotations and correct any mistakes. In the model, we use a wide variety of features derived for a trigger word candidate itself, its immediate context – two words to the left and two words to the right –, as well as the sentence that it occurs in. Hence, some of the variables are the properties of the trigger candidate (e.g. its dictionary form (*lemma*) and part of speech); the properties of its context words (e.g. their lemmas, parts of speech, or whether they belong to any broad semantic category such as locations, names of organizations, human beings); or the properties of the sentence including the trigger candidate, most importantly the paths through the syntactic analysis tree to the nearest animate nouns and place names (to capture information on the location and potential protest actor). We employ *Stanford CoreNLP*, an off-the-shelf NLP toolbox (Manning et al. 2014), to conduct most of the necessary linguistic analyses.

[11] A pairwise averaged F1-score of 0.68. F1-score, a harmonic mean of precision and recall, provides of a single-number summary of the performance of a prediction model (or annotator) and is commonly used when the distribution of the dependent variable is skewed (Manning et al. 2008). F1-score averaged over pairs of annotators is a common way of extending this metric to a multiple-annotator setting (Hripcsak and Rothschild 2005).

In the end, we fit a logistic regression model with the elastic net penalty as we did for the document classification. Also, we tune the classification threshold to obtain a higher precision as described in appendix A. On a test set, this classification model yields a precision of 0.40 and a recall of 0.71. It is easy to use this model for the identification of protest-relevant documents: If the classification model finds no event trigger in a document, then this document gets filtered out. As the model associates probability estimates with all the triggers that it identifies in a document, we take the highest probability that the model assigns to one of the triggers as its estimate of the relevance of that document.

To better understand the performance of the event trigger classification, we manually look through problem cases, which are defined as reports, which the document classifier identifies as relevant but the event trigger classifier as irrelevant. Just like in the evaluation of the document classification model, we look at about three hundred documents for each of Hungary, the UK, and Spain. The correlation in general is high, i.e. if the document classification model finds a document relevant, then the event trigger classification model mostly finds it relevant too. As a result, there are simply not enough problem cases for Hungary in the years 2004 and 2012. We therefore draw on problematic reports from 2004 until 2012 for Spain and the UK and the years 2004 until 2006 and 2013 until 2014 for Hungary.

For the manually evaluation, we draw proportionate samples stratified according to 0.05-probability-intervals that a report is relevant. In other words, for each country, the number of randomly sampled reports in a 0.05-probability-interval is proportionate to the number of the country's reports falling in that interval but scaled down so that the overall sample does not exceed three hundred. A domain expert then manually coded all the sampled documents.

First, we find that it is 2.5 times more likely that a document from the UK is irrelevant than relevant when it is filtered out, and 2.4 times more likely for Hungary. For Spain, however, such a document is about as likely to be irrelevant as it is to be relevant. Next, we compare the distributions of action form types in the documents sampled for the event trigger classification test to those in our dataset (see table 2). Looking at cross-country differences, we note a very high number of blockade events for the UK in the event trigger classification test. For all other action forms and countries, we find similar distributions as in the final dataset. When comparing the last two columns of Table 2, we observe that the distribution of events across the six action forms is similar in the event trigger classification test and the final data. The main difference appears in protest events classified as having another action form. We find more of these types of events in the final data. Again, this is not a major problem as these events are the most difficult to identify and manual coders also tend to disagree on them more.

Table 2: *Comparison of action forms in the event trigger classifier test vs. the final data.*

| | **below 0.85 (event trigger classifier test data)** | | | | | **above 0.85 (final data)** |
|---|---|---|---|---|---|---|
| | **Hungary** | **Spain** | **UK** | **other countries[1]** | **Total** | **all three countries** |
| Demonstrations | 57.8 | 61.5 | 32.4 | 57.7 | 47.6 | 43.2 |
| Violent protest | 31.1 | 34.6 | 22.5 | 26.9 | 27.4 | 25.7 |
| Blockades | 6.7 | | 19.7 | | 10.1 | 8.6 |
| Strikes | 2.2 | 3.9 | 15.5 | 7.7 | 8.9 | 11.1 |
| Petitions | 2.2 | | 8.5 | 7.7 | 5.4 | 7.0 |
| Other protest | | | 1.4 | | 0.6 | 4.4 |
| Total (n) | 45 | 26 | 71 | 26 | 168 | 886 |

[1] Sometimes the event found through manual coding take place in other countries than the one the document was assigned to. In the case of the event trigger classifier test data, 26 events took place in a country that is not Hungary, Spain, or the UK.

*Sampling of protest-relevant documents*

After the six steps discussed so far, the corpus of relevant reports still remained too large to manually annotate all of them. Therefore, we additionally had to sample the documents. This sampling, however, needed to be sensitive to the number of documents available for the different countries. Most notably, we needed to include all documents for the countries for which we already had very few documents, in order to receive at least a minimum of information of contentious mobilization in these countries. For the following countries for which we had a very large sample of documents, we manually annotated 25 percent of the documents: Czech Republic, Germany, Spain, France, UK, Hungary, Ireland, Italy, Latvia and Poland. For countries with an average number of documents (Belgium, Denmark, Estonia, Greece, Norway, Portugal and Slovakia), we sampled 50 percent. Finally, for very small countries with only a few hundred news reports, all the reports were manually annotated. These countries are Austria, Bulgaria, Switzerland, Cyprus, Finland, Island, Lithuania, Luxembourg, Romania, Malta, Netherlands, Romania, Slovenia and Sweden.

## 2. Manual coding of protest events

Our ultimate goal was to retrieve information on all politically motivated unconventional actions in the selected countries and time period. To this end, we employed a simplified version of the PEA as developed by Kriesi et al. (1995) and Hutter (2014) on the filtered and sampled corpus of newswire reports. The full codebook as well as a description of the annotation interface used during the annotations can be found in appendix B.

The first step in the manual coding procedure was the identification of relevant documents, that is, documents reporting on protest events that took place in one of the thirty countries that we study for the years 2000-2015. Already this task is a challenging one (see Ruggeri et al. 2011). The complexity of the task is related to the high degree of interpretation that is required in deciding whether an event is relevant or not. Our coders did not rely on a theoretical definition of relevant protest actions, which might be conceptually precise but practically very difficult to implement. Instead, they identified relevant actions using a detailed list of unconventional or non-institutionalized political action forms. We instructed them to select all mentions of actions listed in the description in *Table 3* as relevant. A relevant document may contain references to only one protest event or it may contain references to more than one. Moreover, the same protest event may be referred to several times in the document. For coders to distinguish events and to annotate them only once, we defined unique events using the combination of the form of action, the timing and the location. Hence, two descriptions of protest events in a document refer to a single event if they happen on the same day in the same community and with the same form of action. If one or more of these key elements differed, the coders annotated multiple events.

Once the relevant documents were identified and the events were singled out, the coders annotated the following variables for each event: 1) the date of the event (the day, month and year when it took place); 2) the location of the event (if possible not only the country but also the city), 3) the form of action, 4) the number of participants, 5) the issue of the protest, and lastly 6) the actors participating in or organizing the protest event (see Table 3 for details).

Table 3: *Form of action, issue and actor categories used in manual coding*

| | | |
|---|---|---|
| **Form of action** | | |
| Demonstrations | demonstrations, marches, rallies, camps, meetings, vigils and other non-confrontational gatherings | |
| Strikes | industrial action of any kind (incl. work stoppages, pickets) | |
| Blockades and related act. | blockades, occupations, sit-ins, camps and other confrontational strategies | |
| Petitions | petitions, letters, consumer activism, boycotts, symbolic protests (performances) | |
| Violent protests | sabotage, riots, destruction of private or public buildings, bomb or arson attacks, violence against persons, clashes with police, cyber-attacks | |
| Other protests | All other action forms | |
| **Issues** | | |
| Economy (private) | Economic claims addressed to firms/ employers: dismissal of staff, closure of firm/branch, labor conflict related to pay rise, pay cut etc. | |
| Economy (public) | Economic claims addressed to public institutions, e.g. welfare, budget policies, agricultural policies, labor regulation, state regulation of the economy more generally | |
| Environment | Environmental protection, nuclear energy, other forms of energy production, infrastructure projects (e.g., transport), animal rights | |
| Cultural liberalism | Peace (questions of war & peace, nuclear and other conventional weapons, military infrastructure, military spending, military service), women's rights (incl. equal treatment, abortion), LGTB (rights and recognition of lesbians, gay, transsexuals, bisexuals), international solidarity (development aid; anti-imperialism), anti-racism, rights of migrants more generally, squatter mobilization (for autonomous living and cultural spaces) | |
| Regionalism | Separatism, regional independence, protection of regional interests, anti-regionalist counter-mobilization | |
| Cultural conservatism | Counter-mobilization to all aspects of cultural liberalism except for anti-racism and migrants' rights (which is in xenophobia) | |
| Xenophobia | Right-wing extremism, racist mobilization (against foreigners or ethnic minorities), anti-immigration | |
| Political | Representation, corruption, electoral reform, and institutional issues in general | |
| Other | All other issues not covered by the previous categories | |
| **Actors** | | |
| Political parties | Left | e.g. Social Democrats, Greens, Communists |
| | Right | e.g. Liberals, Christian Democrats, Conservatives, etc. |
| | Unknown | Unknown or unclear political orientation |
| Trade unions | Public | Public sector unions |
| | Private | Private sector unions |
| | Both | Unions from public and private sectors |
| | Unknown | Unions for which sector is unknown |
| Other organizations | | |
| Social groups | Occupational | Workers, teachers, lawyers, etc. |
| | Students | |
| | Pensioners | |

We collected inter-annotator agreement scores before and during the coding. More precisely, 14 of the 35 coders received the same 65 documents at different times so we measured the level of agreement on both the identification of events and the coding of the additional event attributes. We focused on the most important attributes, which were the actors, issues and number of participants. For the identification of the events, we assessed whether two coders agreed on the date, country and form of action for every event they identified in the same document. Since this is an open-ended identification task – theoretically there is an infinite number of possible date-country-action combinations – we report the unweighted F1 score, which indicates the ratio of matched to non-matched events. The average F1 score over all possible pairs of the 14 coders was 0.60, with a standard deviation of 0.06. As for the event attributes, we first aggregated the matches and non-matches for every attribute of each event which two coders had identified in the same document. Second, we calculated Cohen's Kappa on these aggregated numbers of agreements. We had a fixed number of classes for every attribute, which implied that our measure of agreement had to be chance corrected. The average of Cohen's Kappa for the actors was 0.57 (standard deviation: 0.13), 0.53 (0.13) for the issues and 0.45 (0.06) for the numbers of participants. Although it has been claimed that guidelines on how Cohen's Kappa values should be interpreted are sometimes misleading (Gwet 2014), values from 0.40 to 0.60 are often defined as fair to good.

In spite of all efforts to filter out irrelevant documents during the automated selection of reports, we still ended up with 59.0 percent of irrelevant reports in the manual annotation phase. 26.3 percent of these were false positives that did not contain any events at all. In addition, another 32.6 percent only contained information on events that were already reported in other documents. Overall, 21,173 of the 45,680 documents sampled included unique information on protest events. Hence, our (semi-)automated selection procedure was not able to eliminate the problem of event duplicates in a satisfactory manner.

Yet the manual annotation has its drawbacks, too. In our case, 34 undergraduate and graduate students performed the manual coding. One coder annotated at most 3,092 and at least 417 documents. The amount of time required to annotate a document varied greatly across news agencies and coders. Nonetheless, using the coders' self-declared working time we estimated an average processing time of 7 minutes per document. This estimate provided for a rough assessment of the amount of time required for the coding of the whole corpus of 45,680 documents. We invested about 642 working days in annotating our dataset covering 30 countries over a period of 16 years.

---

## 4. Weighting the data

The main indicator to measure the mobilization of protest is the frequency of events, i.e., the sheer *number of events* in a specific unit (in a region, country etc.) and per a specific time period (per week, month or year). However, nothing is really simple in the study of protest, not even an indicator apparently as simple as the number of events. More precisely, an indicator summing numbers of events per month has inherent features that require special measures before it can be used to compare mobilization in different countries. To begin with, recall that we used separate sampling probabilities in small, medium-sized and large countries. For the numbers to be comparable across countries, we therefore have to adjust for the different sampling probabilities. Accordingly, numbers from countries where we drew a .25 sample were multiplied by 4, and those from countries where we drew a .5 sample were multiplied by 2.

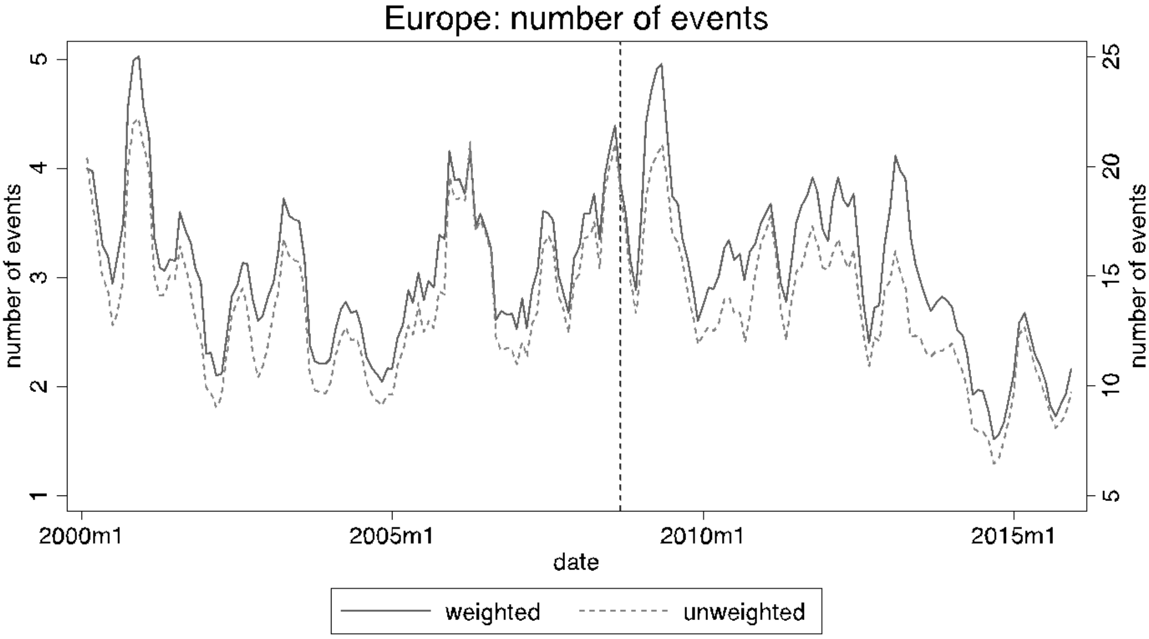
Furthermore, following the example of Beissinger and Sasse (2014), we decided to adjust the number of events according to the country size. The size of the population has important implications for the level of protest we report for a given country. If we do not adjust for country size, we privilege the large countries, i.e., the large countries are the ones that are characterized by the largest numbers of protests. Indeed, if we do not adjust for country size, notoriously unruly France turns out to have the largest number of protests per month (65.8), followed by the United Kingdom (56.9), Spain (52.6), Italy (48.0) and Germany (47.0). We tested different ways of weighting for the country size, and it turned out that a natural logarithmic scale worked best. More precisely, we divided the number of events by the logarithm of the size of the country where the events took place. Accounting for population size *per se*, as Beissinger and Sasse (2014) do, is overcorrecting the data. The large countries are 'punished' too much, and protests in smaller countries become so prominent that all small countries are pushed into the upper half of the average rankings. Weighting with a logarithmic population size scale renders this weight less incisive and the country rankings look more valid. Unfortunately, this adjustment has the drawback that the numbers cannot be straightforwardly interpreted on an intuitive linear scale. However, there is no magic bullet solution to this weighting issue, and after careful thought we considered that the proposed solution offered the best possible compromise.

Lastly, our multisource approach included news agencies focusing on specific countries and we had to correct this focus by using an additional weight. We observed that in our dataset only two small countries have an above average number of monthly protests, namely the Czech Republic (45.3) and Greece (38.2). While Greece is, indeed, a highly unruly country, the high level of protest resulting in the Czech Republic was due to the fact that for this country we relied on a country-specific English-language news agency. It so happens that the same 'advantage' applied to some of the large countries, too. Thus, France

(AFP), the United Kingdom (PA), and Germany (DPA) were the countries which hosted our most important sources, and we also relied on national news agencies for Italy (ANSA) and Poland (PAP), which means that these countries are probably also over-represented in our data. Consequently, we added another weight to control for agency bias. For every country, we distinguished between the share of documents from general news agencies and the share from news agencies specialized in the country. We defined the specialized agencies as follows: ANSA = Italy, AFP = France, DPA = Germany, BNS = Baltic states, MTI = Hungary, PA = UK, Ireland and Northern Ireland, BBC = UK and Northern Ireland, PAP = Poland and CTK = Czech Republic. To construct the agency weight, we took a mean of 1 and the general news agency share. If all the documents are from a specialized agency, the weight is 0.5; if all are from a general agency, the weight is 1.

*Figure 3* presents the ups and downs of protest across European countries with the weighted and unweighted data. It shows that the weighting procedures implemented for the country comparisons do not affect the overall trend in protest. During the 16 years that we analyze, the two lines capturing fluctuations in the amount of protest move together. Importantly, the peaks correspond in both datasets. The main difference lies in the fact that the weighted data accentuate some of these peaks. As an additional control, we calculated the correlation between the two patterns. It is quite close: r=.89 for the smoothed curves presented in *Figure 3*, and .90 for the corresponding raw curves.

Figure 3: *Number of events across Europe, 2000-2015: moving averages, weighted and unweighted*




## 5. External validation and aggregating the data

In this his section, we compare our data set to two comparable existing data sets in order to establish the external validity of our approach. The comparisons will additionally show that the aggregation of the protest event data to the country-month level as well as using 5-monthly moving averages are the most meaningful ways to pool our data.

The first comparison is made with the Integrated Crisis Early Warning System (ICEWS) data (Boschee et. al. 2015). Among other things, the ICEWS data contain information on protest events for all the 30 countries and almost all the years covered by our study. This enables a detailed look into the differences across countries. More precisely, we can assess whether specific unwanted country characteristics such as the size of the country introduce a bias into our analyses. The ICEWS data are only available until 2014, whereas our time period extends to 2015. As for the substantial information, the ICEWS data only contain information on the form of action, and only the definition of the demonstration[12] is comparable to our definition. In the ICEWS annotations, events are derived from both multiple English news sources and foreign language sources (Spanish, Portuguese and French), which are machine translated into English prior to the analysis. Compared to our data, the ICEWS data are therefore built from similar types of sources, but the number of sources is larger and the set includes sources in languages other than English as well. Compared to our semi-automated approach, in the ICEWS data set the event detection is performed fully automatically. More precisely, the news reports are processed using the commercial software BBN ACCENT, which employs a range of natural language tasks and dictionary lookups in the automated recognition of events.

The second comparison of our semi-automated data set is with data derived manually from national newspaper articles and news reports for a selection of western, eastern and southern European countries (Spain, Germany, the Netherlands, the UK, Hungary and Poland). These data sets were collected by different teams of researchers and therefore cover different periods of time. Nevertheless, they allow us to compare the numbers and the types of events that we find in the news reports published in English to the national press. This allows us to estimate potential biases related to the relative sparsity of our data. Such a comparison can establish which kinds of events make it into the international press, and which events are systematically excluded.

**Comparison with the ICEWS data**

The validation of our data with the ICEWS data will be conducted as follows. In a first step, we establish the general trends in the data sets concerned and examine whether

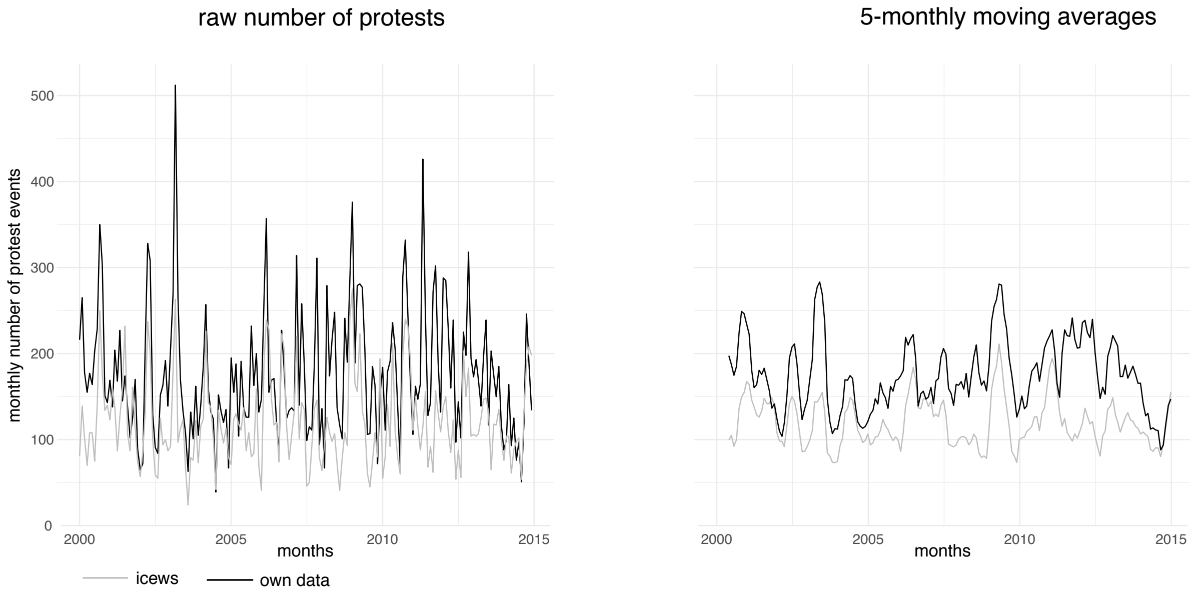
[12] In both data sets, demonstrations are defined as politically motivated public gatherings such as demonstrations, marches, rallies, camps, meetings, vigils etc.

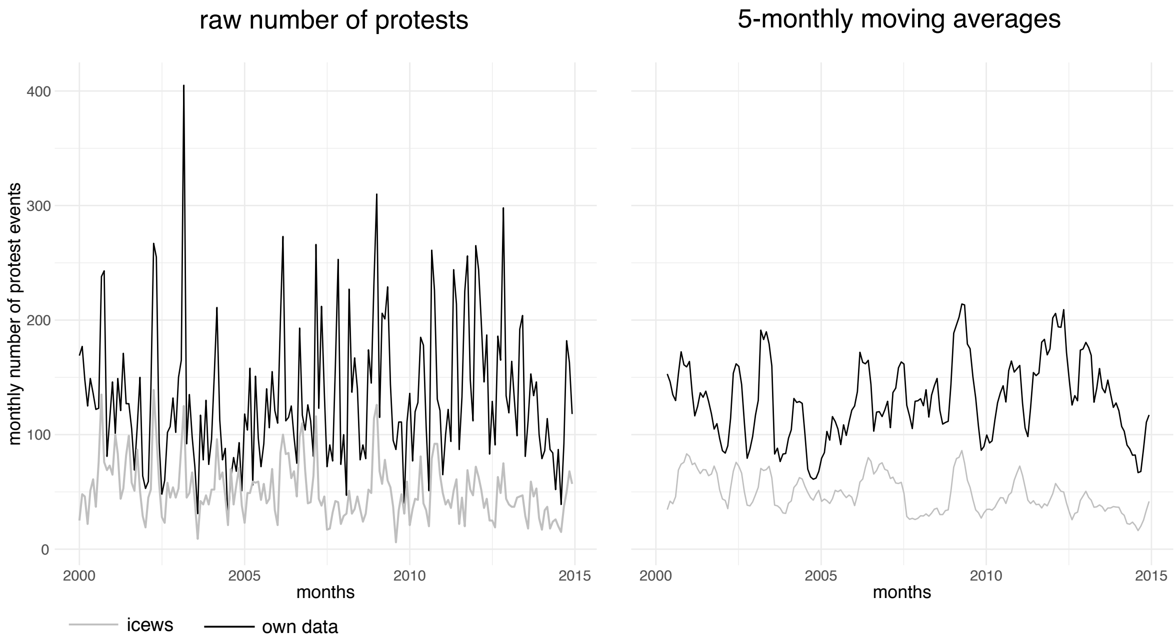
different weighted averages are beneficial to the validity of the data. In a second step, we show that the data sets compare very differently across the 30 countries and that general country characteristics cannot explain the reasons for these differences. Finally, in a third step we present an in-depth qualitative analysis of the time periods for which the two data sets deviate most heavily in terms of the numbers of events they report.

All the comparisons will be conducted in two ways. First, all the events annotated as demonstrations in the two data sets are compared. This allows an assessment of the general differences between the two data sets. For the second comparison, both data sets are restricted to events collected from sources that are available in both data sets. More precisely, only events from BBC Monitoring, Agence France-Presse (AFP), Associated Press (AP), the Czeck News Agency (CTK), the Baltic News Service (BNS) and Agenzia Nazionale Stampa Associata (ANSA) are included.

The most important question, and the one with which we therefore start, is whether the overall levels and the dynamics of the time series are similar. With this aim, the trends in the two time series over all the countries are plotted in Figure 4. On the left-hand side, the monthly aggregated numbers of protests are shown, while on the right the 5-monthly moving averages are shown.

Figure 4: *Monthly number of protest events in the full (top) and restricted (bottom) data sets for the ICEWS and our own protest data*



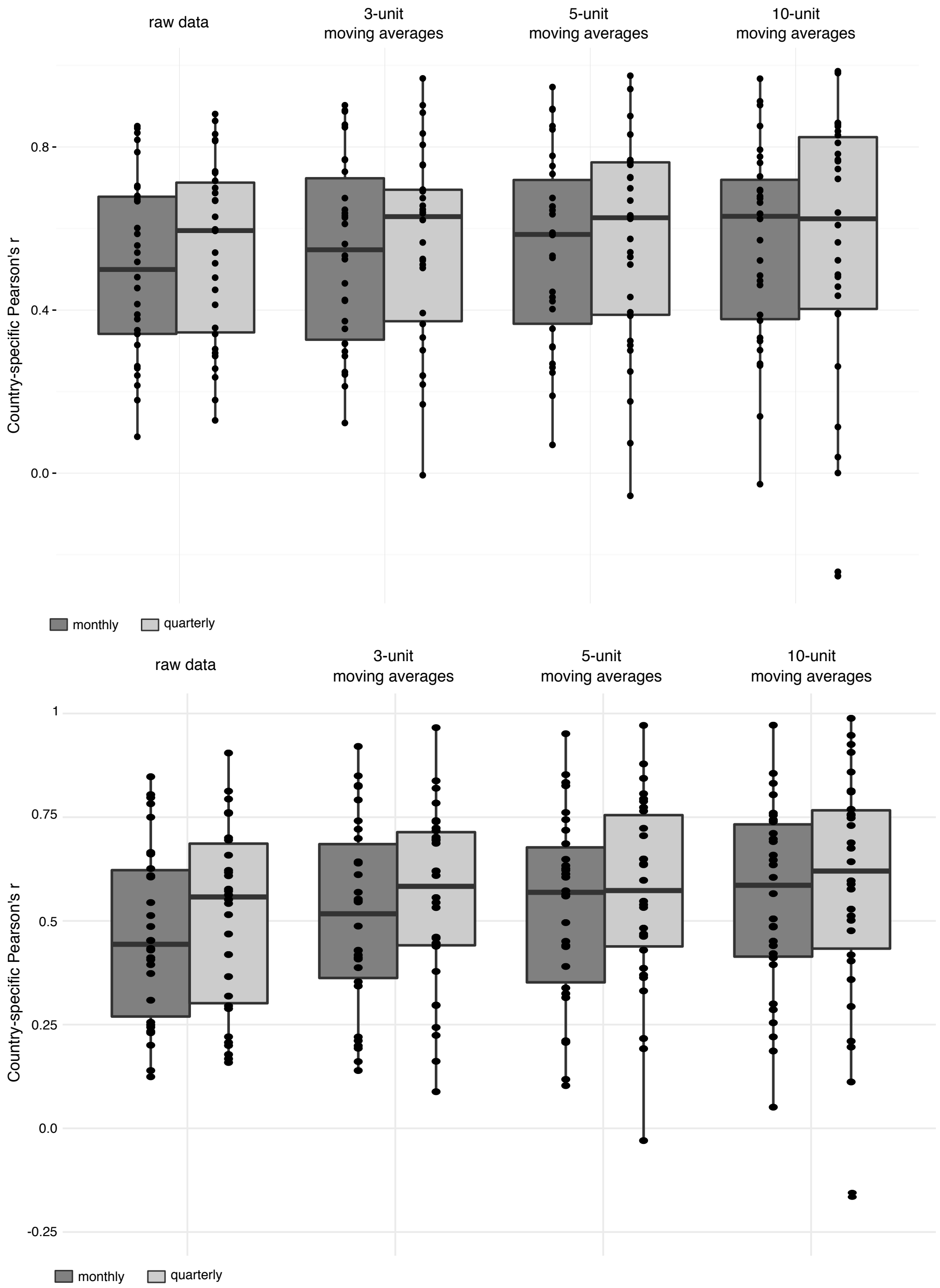




First, it is clear that the level of the full ICEWS time series is far lower than in our data. There can be different reasons for this. Our data may significantly over-report the numbers of protests or the ICEWS data might under-estimate the number of protests. Later on, we will present the results of a qualitative evaluation of these differences that will show that the latter is most likely the case. Most notably, the ICEWS data even contain many duplicated events, and violent demonstrations, which, according to the ICEWS codebook, should have been categorized as a different form of action (violent protests instead of demonstrations). The restricted data sets, in contrast, differ much less. For many months, both data sets report similar levels of protest. Secondly, the ICEWS data show regular peaks, while our data only show three substantial ones: the Iraq war demonstrations in 2003, the protests following the meltdown of the global financial system in 2009, and the protests during the Eurozone crisis in the years after 2010. For the restricted data set, these patterns are less pronounced. Nevertheless, our data still seem to accurately show the three major peaks, while they are again less visible in the restricted ICEWS data set. Finally, a comparison of the raw numbers on the left and the 5-monthly moving averages on the right shows that for descriptive and visual analyses the data should be standardized to a certain extent.

So far, it has been shown that the differences over time are constant and thus most likely not a concern for the external validation of our data compared to the ICEWS. In contrast, however, in the following we show that the external validity indeed heavily differs depending on the single countries. Figure 5 shows the distribution of correlation coefficients for the numbers of protests across the 30 countries under consideration for different levels of aggregation and varying strengths of smoothing. *On average*, we can achieve a better congruence between the two data sets for both a higher level of aggregation and more smoothing. If we compare quarterly time series instead of monthly time series, the median for the correlation coefficients increases from 0.50 to 0.54 for the full data sets (at the top of Figure 5). If we additionally increase the level of smoothing, the median correlation coefficient increases to a maximum of 0.56 for the time series with 5-quarterly moving averages. By increasing the smoothing and the level of aggregation, we can therefore assume that some of the annotation and selection errors are straightened out and the data sets become more comparable. The correlations for the restricted data sets behave very similarly. Quite surprisingly, however, they are generally lower, despite the fact that the two data sets start from the same sources. For the majority of the countries, the correlation coefficients are weak, i.e., below 0.6. However, they are still strong for about a third of the countries.

Nevertheless, while the average congruence increases, so does the *spread* among the countries. A comparison of the time series with 10-quarterly moving averages is the most extreme. While the two data sets correlate with an impressive 0.98 for Greece, the correlation is clearly negative for Finland and the Czech Republic (-0.24 and -0.25 respectively). Hence, we face a *trade-off between the average level and the variation* in the two data sets. The comparison of the monthly aggregated time series smoothed with a 5-monthly average seems to offer the best solution to this trade-off. The correlation for Finland is still virtually non-existent (0.07), but besides this singular odd result, no other country has a lower correlation coefficient than 0.20.

Figure 5: *Country-specific correlations between the ICEWS and our own data for different levels of aggregation (monthly, quarterly) and different types of moving averages (3,5 or 10). Comparisons between the full data sets are shown at the top, comparisons between the restricted data sets are shown at the bottom.*



The variation in the difference between the two data sets seems to systematically correlate with general country characteristics. Table 4 illustrates this by showing the results of regressions on the difference in the 5-monthy averaged times series for the full and restricted data. The dependent variable is the monthly difference between the ICEWS data and our data set.[13] As explanatory factors, we include the regions (Central Eastern, Northwestern and Southern Europe), the size of the population (in millions), a binary indicator of whether a country in our sample is a self-declared focus country for one of the news agencies in our sample, and three of our four time periods (normal times, the shock period and the Euro-crisis period. We show the results of a random effects model in order to account for uneven variance across months and countries.

Table 4: *Random-effects GLS regression on the difference between 5-monthy averaged ICEWS and our own data*

| | ***Full data set*** | | | ***Restricted data set*** | | |
|---|---|---|---|---|---|---|
| | *Coef.* | *Std. Err.* | *P>\|z\|* | *Coef.* | *Std. Err.* | *P>\|z\|* |
| (Intercept) | -0.255 | 0.820 | | -0.064 | 1.017 | |
| Region (ref = central eastern) | | | | | | |
| Northwest | 0.860 | 0.974 | | 0.412 | 1.205 | |
| Southern | -0.576 | 1.227 | | -2.393 | 1.519 | |
| Average population (in m. inhabitants) | -0.050 | 0.021 | * | -0.049 | 0.025 | * |
| Agency focus | -2.613 | 1.004 | ** | -3.369 | 1.237 | ** |
| Period (ref = normal) | | | | | | |
| Shock | -0.789 | 0.205 | *** | -0.885 | 0.171 | *** |
| Euro-crisis | -0.264 | 0.125 | * | -0.908 | 0.105 | *** |
| N observations | 5,280 | | | 5,280 | | |
| N countries | 30 | | | 30 | | |
| Wald test | 43.64 | DF = 6 | *** | 114.71 | DF = 6 | *** |
| R-squared | within: | 0.00 | | within: | 0.02 | |
| | between: | 0.54 | | between: | 0.52 | |
| | overall: | 0.19 | | overall: | 0.27 | |

*Notes:* Unstandardized GLS coefficients and standard error.
Signifiance codes: *** p ≤ 0.001, ** p ≤ 0.01, * p ≤ 0.05.

First, the size of the differences does not depend on the region. Compared to the reference category, central eastern Europe, the differences are non-significant for northwestern and southern Europe. Furthermore, the country size is negatively correlated

[13] Since the data are stationary, no transformations such as de-meaning or de-trending are necessary.

with the differences between the two data sets. This is not surprising, since it can be assumed that international news agencies are more present in larger countries, which most likely increases the number of protest events indicated in both data sets, and thus potentially leads to smaller differences. In addition, the agency focus is related to smaller differences. There is most probably some under-reporting for some countries which are not in the spotlight of our international news agencies. Hence, we have most likely introduced some bias into the data, which is why we develop a specific weight for this (see section 4 on weights). Finally, there is also a relationship between the periodicity and the differences in the two data sets. The differences in the shock and Euro-crisis periods are substantially smaller than before the Great Recession. This can be explained by the fact that the crisis years in general might be better covered by international news agencies – more public attention is given to protests, to which news agencies have to cater. In consequence, the selection bias in the sources of both data sets is smaller, which should lead to smaller differences between the time series.

*Qualitative comparison*

The general comparison of the two time series so far has revealed little about the actual reasons for the differences. It is, however, crucial to know more about the substantive differences in the two data sets. Otherwise, we cannot confidently start with our analysis, especially for the countries which show a low overall correlation. To shed more light on the substantial differences, we conducted a qualitative examination of the time periods where the deviations between the two time series are most pronounced. More precisely, we compiled our set of problematic time periods for this check as follows. First, we looked at the highest monthly differences per country in the raw data between the ICEWS and our own data. From these periods with the highest monthly differences, we chose a case from one country in each of the four regions. For these cases (see Table 5), we calculated the duplicates in terms of location and date and tried to assess the difference in the number of protests by means of a study of country reports, yearbooks, newspaper articles, Wikipedia entries and general web searches.

Table 5: *Selected cases for the qualitative evaluation*

| Month | Country | N protests own data | N protests ICEWS | Difference | Duplicated events ICEWS |
|---|---|---|---|---|---|
| **Dec. 2009** | Denmark | 6 | 86 | -80 | 70 |
| **Sept. 2006** | Hungary | 34 | 153 | -119 | 131 |
| **Dec. 2014** | Ireland | 2 | 28 | -26 | 18 |
| **July 2001** | Italy | 14 | 97 | -83 | 75 |
| **March 2003** | Ireland | 13 | 3 | +10 | 0 |
| **March 2007** | Norway | 4 | 0 | +4 | 0 |
| **Jan. 2012** | Romania | 74 | 56 | +18 | 36 |
| **May 2011** | Portugal | 11 | 2 | +9 | 1 |

In the first case, Denmark in December 2009, our search through external sources only found a few demonstrations, as our data set reports. We also found many events that we included in other categories, such as blockades, petitions and violent protests. Such protests should have been included in other categories in the ICEWS data as well. However, the even bigger reason for the differences is duplicated events. While we systematically filter them out in our data set, the ICEWS contains 70 in this case. Such demonstrations that take place on the same day in the same place as other already annotated demonstrations make up the lion's share of the difference. The same holds true for Hungary in September 2006, where many duplicates can be found and many violent protests took place. The third case, Ireland in December 2014, was less clear. There are many duplicates here too, but there remains a substantial amount of over-reporting in the ICEWS data. Even after a very extensive search, we could not find more than the 2 events reported in our data set. These are related to the 'right 2 water' protests in Dublin. For Italy in July 2001, we also did not find more events than are reported in our data. There is much reporting on police violence related to the events mentioned, so maybe these are falsely counted as separate events by the ICEWS. As for the four cases in which our data set reports more events than the ICEWS, we found most of the events that are reported in our data set with the exception of Portugal in May 2011. March 2011 was a major protest month in Portugal (organized by the 'Movimento 12 de Março'), so we suspect that our coders had problems finding the correct dates related to these protests. So, in these cases, the ICEWS seems to under-report protests. However, compared to the cases where the ICEWS data report more events, these cases are rare and the differences are not very large.

Hence, this qualitative assessment of the largest differences between the two data sets yielded the impression that the ICEWS data contain duplicates and events that would have been classified in other categories by our coders. The ICEWS data are compiled from many more news sources than our data set, which most probably also leads to many more duplicated entries of the same protest event. Sometimes, however, events such as police violence, which should not be counted as protests, are reported by the ICEWS.

**Comparison with data based on national news sources**

In this second part, we compare our protest event data to data sets based on the annotation of national news sources, all published in the relative national language.[14] For Spain, the source is El Pais and the data cover the years 2007 to 2015.[15] For the west European countries, the sources are quality newspapers (for the UK, The Guardian; for Germany, Frankfurter Rundschau and for the Netherlands, NRC/Handelsblad) and the dataset is based on annotations of Monday editions only. For the northwest, we have data for the period running from 2000 to 2011.[16] For central and eastern Europe, finally, the data cover the same period

[14] The national data sets cover different periods of time: our whole period of interest (for the CEE countries), finishing before the end of our period (northwest), or not starting at the beginning of the 2000s (south). In order to account for these variations, we include in the comparison only data for the same time periods as are available in the national news.

[15] We thank Martin Portos for sharing his protest data set.

[16] We are grateful that Swen Hutter provided the protest data on the Northwestern countries to us.

as ours – 2000-2015 – and the sources are national news agencies: MTI (Hungarian News Agency Corporation) for Hungary and PAP (Polish Agency Press) for Poland.[17]

Following what we did in the comparison with ICEWS, in the comparisons below we only weight the data according to the sampling probability used to select documents annotated for the individual countries.

*Forms of action and news sources*

We start our comparison with the share of different forms of action found in the data sets based on the annotations of international and national news sources. This comparison is important for our subsequent analyses, in particular for the study of action repertoires and how they evolve during economic and political crises. Table 6 shows that demonstrations are the most common type of protest events across the data sets and countries. For four countries, demonstrations represent more than half of all events, while for Spain and the Netherlands they amount to 40 percent of all events. Concerning the share of demonstrations across news sources, we find the same share for three countries (Hungary, Spain and the Netherlands). For Poland and the UK, the share of demonstrations is greater in the national news, while for Germany it is greater in the international news.

[17] We like to thank Ondrej Cisar for giving us the protest data that he collected for Central and Eastern Europe.

Table 6: *Forms of action by news source for the six countries*

| | | | Demo. | Strike[2] | Violent | Other | N[3] |
|---|---|---|---|---|---|---|---|
| Northwest | Germany | International | **60.5** | **-** | **31.2** | 8.3.0 | 5,908 |
| | | National | 48.6 | - | 19.7 | **31.7** | 1,697 |
| | Netherlands | International | 44.2 | - | 42.2 | 13.7 | 351 |
| | | National | 41.6 | - | 36.4 | **22.0** | 341 |
| | UK | International | 53.2 | - | **31.5** | 15.3 | 6,988 |
| | | National | **60.1** | - | 22.1 | 17.8 | 449 |
| Central and East | Hungary | International | 50.0 | **19.2** | 19.2 | **11.6** | 2,104 |
| | | National | 52.6 | 5.1 | **33.9** | 8.5 | 3,716 |
| | Poland | International | 48.0 | **15.4** | **20.6** | 16.0 | 3,376 |
| | | National | **61.0** | 6.4 | 14.6 | 18.0 | 6,713 |
| South | Spain | International | 44.8 | **14.1** | **30.1** | 11.0 | 5,404 |
| | | National | 45.0 | 10.4 | 19.8 | **24.8** | 2,002 |

*Notes:*
[1] Bold cells have higher adjusted residuals than expected (>1.96).
[2] Strike data are not available for the northwestern national news datasets.
[3] The total number of events is multiplied by the sampling probability. We multiply the events found for Germany, the UK, Hungary, Poland and Spain by four. For the Netherlands, we coded all the documents found so the total number of events is not corrected for sampling probability.

The second form of action we are interested in is strikes. Unfortunately, we do not have data on strikes for the northwestern countries so we exclude them from all our comparisons between national and international news for the northwest. We notice that strikes account for a smaller share of all events, less than 20 percent. Regarding the comparison of interest to us, that between national and international news, strikes account for a larger share of the events in the international news. For Hungary, there are four times as many strikes in the international news (19 percent) as in the national (5 percent). Similarly, for Poland the ratio is one to three, with 15 as opposed to 6 percent). Lastly, in Spain the difference is smaller (14 vs. 10 percent) but also biased in favor of international news.

Surprisingly, violent events account for a larger share of events than strikes. They represent between a third and a fifth of all events depending on the type of news source and the country. In the case of violent events, as we found for strikes, they tend to be over-reported in the international news. This is the case across regions, for Germany, the UK, Poland and Spain. For the Netherlands, no difference across sources is observed, while for Hungary violent events are over-reported in the national news. Less surprising is over-reporting of other action forms – the residual category – in the national news compared to the international news. National newspapers have more space to report on other, perhaps more innovative or less conventional, forms of action than do international news sources.

Our comparison of forms of action across news sources reveals no unexpected biases in the way international news sources report on protest events. They tend to over-report protest events deemed newsworthy – in particular, violent events. The newsworthiness of a protest event is set at a higher level in the international news than in the national. Therefore, events including violence are more likely to be reported. This appears for four of the countries we used in the test and we should therefore be careful when analyzing processes of radicalization in the next steps of our analyses. More surprisingly, the international news over-reports strikes. This might be related to the fact that during the crisis economic protest events were a core focus of attention in the international press, more so than in the national news. We will keep this bias in mind when comparing pre- and post-crisis periods, but for cross-national comparison it does not appear to be a problem since we find a similar bias in the two regions for which we have data, the south and the east.

*Location and size of protest events*

Next, we turn to specific characteristics of protest events that might explain why they are reported in the international press. In particular, the location and the size of events are key variables to explore (see Table 7). We see that the international news agencies are biased in favor of events taking place in the capitals, in the whole country, and large events with more than 10,000 participants.

First, regarding the location of the events, we find that the international news is more likely to feature protest events taking place in capitals. This finding holds across all the countries but there are variations in the amount of over-reporting of protest organized in capitals. For Germany, the percentage of events in the capital is twice as large in the international as in the national news. For Hungary, Poland and the UK, the share is 50 percent higher in the international news than in the national. For Spain and the Netherlands, the differences are smaller. When considering large cities instead of focusing on the capital, we find a more diverse picture. In this case, sometimes we find that events taking place in large cities are more likely to be reported in the national news. This is the case for Germany, Hungary and Spain, while for the Netherlands and the UK the share of events taking place in large cities is higher when using the international press than the national.

Table 7: *Location and size by news source for the six countries*

| | | | **Capital** | **Large cities[1]** | **Nation-wide[2]** | **N[3]** | **Large events (>10,000)** | **N[3]** |
|---|---|---|---|---|---|---|---|---|
| Northwest | Germany | International | **23.0** | 22.5 | 11.2 | 5,908 | 19.72 | 3,576 |
| | | National | 11.6 | **33.1** | **18.2** | 1,697 | 12.2 | 1,238 |
| | Netherlands | International | **31.9** | **29.3** | **23.1** | 351 | 16 | 187 |
| | | National | 23.5 | 20.8 | 7.6 | 341 | 3.8 | 211 |
| | UK | International | **47.2** | **11.3** | **15.2** | 6,988 | 13.2 | 3,360 |
| | | National | 32.7 | 7.8 | 7.6 | 449 | 21 | 224 |
| Central and East | Hungary | International | **57.6** | 4.2 | **28.3** | 2,104 | 17.9 | 896 |
| | | National | 36.6 | **8.5** | 3.3 | 3,716 | 4.8 | 1,770 |
| | Poland | International | **40.2** | 13.6 | **17.9** | 3,376 | 14.4 | 1,472 |
| | | National | 25.9 | 15.3 | 7.2 | 6,713 | 2.5 | 4'664 |
| Southern | Spain | International | **33.9** | 16.0 | **22.3** | 5,404 | 25.2 | 2,236 |
| | | National | 27.3 | **31.1** | 0.1 | 2,002 | 16.1 | 1,363 |

*Notes:*

Bold cells have higher adjusted residuals than expected (>1.96). Underscores indicate a t-test confirming that the two samples have different means.

[1] Large cities exclude the capital and comprise cities > 500,000 inhabitants for Germany (Hamburg, Munich, Cologne, Frankfurt am Main, Essen, Stuttgart, Dortmund, Dusseldorf, and Bremen), for Poland (Krakow, Lodz, Wroklaw, Poznan, and Gdansk), and for Spain (Barcelona, Valencia, Sevilla, Zaragoza, and Malaga); > 300,000 inhabitants for the Netherlands (Rotterdam, Utrecht, and The Hague) and the UK (Birmingham, Liverpool, Leeds, Sheffield, Bristol, Manchester, Leicester, Glasgow, and Edinburgh); and > 100,000 inhabitants for Hungary (Miskolc, Nyiregyhaza, Debrecen, Keckskemet, Gyor, Szeged).

[2] For northwestern and southern Europe, the data resulting from annotation of national news do not include a specific code for protest events taking place in the whole country. We create a proxy using a code to identify events taking place on the same date, addressing the same issues, but taking place in different locations.

[3] The total number of events is multiplied by the sampling probability. We multiply the events found for Germany, the UK, Hungary, Poland and Spain by four. For the Netherlands, we coded all the documents found so the total number of events is not corrected for sampling probability.

Apart from these two types of location, it is interesting to look at the share of events taking place throughout the country – which are also more likely to gain the attention of the international press. In fact, in this case they are consistently more likely to be featured in the international press than in the national. Furthermore, the differences can be very large but may also be due to the lack of a clear code for nationwide events in the northwestern and southern national data sets.

Looking at the size of events, we investigate the share of events with more than ten thousand participants reported in both news sources. Large events represent a larger share of all events in the international press than in the national press. This difference is greatest for small countries and CEE countries. The Netherlands, Hungary and Poland display larger differences across news sources with regard to the share of events attracting large crowds. In these cases, the percentages of large events are four times larger in the international news. The differences are much smaller for Germany and Spain, and virtually absent for the UK. It is not surprising that the international news focuses on big events, especially those taking place in specific countries that are not of focal interest to international audiences.

*Issues and actors in protest events*

Lastly, we look at issues and actors to identify biases in specific issues or actors (see Table 8). For the issue addressed by protesters, we compare events focusing on economic issues, and for the CEE countries also on issues related to democracy (for the other regions, this was not included in the list of issues annotated). For the economy, we find important variations across countries. However, there is no prevailing trend in the over- or under-reporting of economy-related protest in the international press. In Germany and the UK, the international press reports a larger share of events focusing on the economy than does the national press. However, the contrary is true for Hungary and Spain. For the Netherlands and Poland, there are no differences with regard to the prevalence of economic issues across news sources. For issues related to democracy in the CEE countries, they tend to be over-reported in the international press.

Table 8: *Issues and actors by news source for the six countries*

| | | | Economy | Democracy[1] | Parties | Trade unions | CSO | N[2] |
|---|---|---|---|---|---|---|---|---|
| Northwest | Germany | International | **17.7** | 12.9 | 10.4 | 7.9 | 16.1 | 5,908 |
| | | National | 6.1 | - | **15.8** | 7.7 | **28.7** | 1,697 |
| | Netherlands | International | 16.5 | 15.6 | 3.1 | 4.0 | 19.9 | 351 |
| | | National | 12.6 | - | **11.4** | 6.8 | **27.9** | 341 |
| | UK | International | **26.5** | 8.8 | 3.5 | 9.9 | 27.0 | 6,988 |
| | | National | 10.0 | - | **6.9** | 8.5 | 31.6 | 449 |
| Central and East | Hungary | International | 42.0 | **29.3** | 12.2 | 17.9 | 13.5 | 2,104 |
| | | National | **61.4** | 16.7 | 11.7 | 15.7 | 12.5 | 3,716 |
| | Poland | International | 51.9 | **18.4** | 8.8 | 36.3 | 19.9 | 3,376 |
| | | National | 52.4 | 4.3 | 9.1 | **42.8** | 17.4 | 6,713 |
| South | Spain | International | 50.4 | 11.1 | 3.0 | 12.4 | 21.3 | 5,404 |
| | | National | **70.1** | - | **10.6** | **51.2** | **48.9** | 2,002 |

*Notes:*

Bold cells have higher adjusted residuals than expected (>1.96) and underscore indicates adjusted residuals smaller than expected (< -1.96).

[1] The issue "democracy" was not available in the northwestern and southern national news datasets.

[2] The total number of events is multiplied by the sampling probability. We multiply the events found for Germany, the UK, Hungary, Poland and Spain by four. For the Netherlands, we coded all the documents found so the total number of events is not corrected for sampling probability.

For the actors, a comparison shows that the composition is very similar for the events reported in the national and the international press. There is an overall tendency in the national press to include more information about the actors engaging in a protest. Furthermore, we find that the reported share of events organized by parties tends to be larger in the datasets based on national sources in the three northwestern European countries and Spain but not for the two central eastern European countries. For trade unions, the frequency is also higher in the national press than in the international, but only for Poland and Spain. For civil society organizations (CSOs), a similar tendency appears for Germany, the Netherlands and Spain. Therefore, the national news often conveys more information about the presence of parties and to some extent also trade unions and CSOs across regions and countries. However, the overall composition – the share of actors of each kind – is consistent across the two types of sources.

## References


ACE. 2005. Ace (Automatic Content Extraction) English Annotation Guidelines.

Beissinger, Mark R. and Gwendolyn Sasse 2014. An End to 'Patience'? The Great Recession and Economic Protest in Eastern Europe, pp. 334-370 in Mass Politics in Tough Times, edited by Nancy Bermeo and Larry M. Bartels, Oxford U Press.

Boschee, E., Lautenschlager, J., O'Brien, S., Shellman, S., Starz, J., & Ward, M., 2015. ICEWS coded event data. Harvard Dataverse, 5.

Friedman, J., Hastie, T., & Tibshirani, R. 2001. The elements of statistical learning. Vol. 1. Berlin: Springer series in statistics.

Grimmer, J., & Stewart, B. M. 2013. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. Political analysis 21(3): 267-297.

Gwet, K. L. 2014. Handbook of Inter-Rater Reliability. Gaithersburg: Advanced Analytics.

Hripcsak, G., & Rothschild, A. S. 2005. Agreement, the f-measure, and reliability in information retrieval. Journal of the American Medical Informatics Association 12.3: 296-298.

Hutter, S. 2014. Protest Event Analysis and Its Offspring, pp. 335-367 in *Methodological Practices in Social Movement Research*, edited by Donatella della Porta, Oxford: Oxford University Press.

Hutter, S. 2014a. *Protesting culture and economics in Western Europe. New cleavages in left and right politics*. Minneapolis : University of Minnesota Press.

Jenkins, J. C. & Maher, T. V. 2016. What Should We Do About Source Selection in Event Data? Challenges, Progress, and Possible Solutions. International Journal of Sociology 46(1): 42-57.

Kriesi, H., R. Koopmans, J. W. Duyvendak, and M. G. Giugni 1995. New Social Movements in Western Europe. A Comparative Analysis. Minneapolis: Univesity of Minnesota Press.

Manning, C. D., Raghavan, P., & Schütze, H. 2008. Introduction to information retrieval. Vol. 1. No. 1. Cambridge: Cambridge University Press.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J.R., Bethard, S., & McClosky, D. 2014. The Stanford Corenlp Natural Language Processing Toolkit. Pp. 55-60 in ACL (System Demonstrations). Baltimore, Maryland USA.

Nardulli, P. F., Althaus, S. L., & Hayes, M. 2015. A Progressive Supervised-Learning Approach to Generating Rich Civil Strife Data. Sociological Methodology 45(1): 148-83.

Pedregosa, F., et al. 2011. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research: 2825-2830.

Ruggeri, A., Gizelis, T.-I., & Dorussen, H. 2011. Events Data as Bismarck's Sausages? Intercoder Reliability, Coders' Selection, and Data Quality. International Interactions 37(3): 340-61.

Sebastiani, F. 2002. Machine Learning in Automated Text Categorization. ACM computing surveys (CSUR) 34(1): 1-47.

Theobald, M., Siddharth, J., & Paepcke, A. 2008. Spotsigs: Robust and Efficient near Duplicate Detection in Large Web Collections. in 31st annual international ACM SIGIR conference on Research and development in information retrieval, edited by ACM.

Zou, H. & Hastie T. 2005. Regularization and Variable Selection Via the Elastic Net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67(2): 301-20.

## Appendix A: Tuning of classification model thresholds

In a binary logistic regression, one predicts the positive class for a data sample $x$ if the estimate of the conditional probability of the positive class given $x$ is greater than 0.5. However, one is free to change the threshold of 0.5 if an application calls for a higher precision (in which case one would increase the threshold) or a higher recall (in which case one would lower the threshold). We tune both classification models to a higher precision as we estimate that after this classification step, we are still left with an impractically large amount of documents for manual coding. For this purpose, we use a set of about 14,000 documents identified by the document classification model as relevant and then checked manually. We experiment with the ensemble of the document and event trigger classification models. The ensemble finds a document relevant if both classification models find it relevant, otherwise the document is declared irrelevant. We tune the ensemble so that it achieves a higher precision and recall on this dataset than the document classification model alone. We keep this setting for the final filtering of documents for manual coding. The document classification model passes through only fourteen percent of input documents, and the event trigger classification model seventy percent. Altogether, this amounts to just above one hundred thousand documents classified automatically as relevant.

## Appendix B: Codebook of manual content analysis

This codebook contains the instructions on the protest event analysis PEA30sixteeen. The first section defines which information in the news wires we read is to be considered relevant in terms of this protest event analysis. The second section details how different references to the same protest event have to be linked. The third section lists and explains the indicators that have to be annotated with respect to every single protest event. Finally, in the fourth section, the annotation interface is described and precise coding instructions as well as examples are specified.

***PLEASE NOTE:***

There are always questions and insecurities during an annotation. Please do not hesitate to ask <name of responsible> in all these cases. It is better to ask many questions than to put the data collection into danger with unassertive annotations!

**CONTENTS**



### 1. RELEVANCE OF PROTESTS

Which documents and sentences have to be annotated?

The decision on whether a document or sentence is relevant for this project depends on whether a relevant protest event is covered in the document. We define a relevant protest event on the basis of three types of information: *location, time* and *action form*.

#### Location

A protest event is only relevant for us if it takes place in one of the following countries we are interested in (all EU countries minus Croatia plus Switzerland, Norway and Iceland):

*Austria, Belgium, Bulgaria, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland, United Kingdom*

In most cases, there is an explicit link between the reference of a protest and a location, e.g. “Thousands demonstrated yesterday in Madrid…” or “The Helsinki protests led to…”, sometimes, however, the location can be inferred from persons or organizations.

**Time**

A protest event is only relevant for us if it happened in the years between

*January 1, 2005, and December 31, 2014*

Mostly, a reference to a date (such as the year, month or even day) or date-like information (such as “yesterday”, “last week” etc.) is explicitly linked to a protest event. In the cases in which such information is not present throughout the whole document, the date of publication of the article has to be assumed to be the date of the protest.

**Form of action**

Our conceptual goal is to collect data on all politically motivated, unconventional actions in the selected countries and time period. In practice, we do not rely on a precise theoretical definition of a relevant protest action but on a detailed list of unconventional or non-institutionalized political action forms:

| **Action form** | **Additional explanation** |
|---|---|
| Strike | industrial action of any kind (incl. work stoppages, pickets) |
| Collections of signatures and petitions | incl. a collectively signed letter or collective letter with is send individually |
| Demonstrations and marches | protesters take to the streets, they move |
| Protest camps/meetings/ vigils/rallies and other festive forms | people do not move, the protest is staged in one place |
| Symbolic protest actions | e.g. street theatre, performances |
| Boycotts | incl. consumer boycotts, school boycotts |
| Cyber-attacks | incl. mail bombings, hacking |
| Hunger strikes | |
| Refusals of payment | loans, taxes, etc. |
| Blockades | incl. sit-ins and picket lines |
| Squattings | Occupation of land or houses |
| Bomb threats | |
| Symbolic violence against objects or persons | e.g. paint bombs, tomatoes, eggs, burning books |
| Other confrontational actions | e.g. mutiny of prisoners, shouting at the president or at any authoritative figure, threats against persons, whistle-blowing |
| Sabotage | |
| Riots | |

| Action form | Additional explanation |
|---|---|
| Destruction of private or public buildings | breaking windows, break-ins, burglary, etc. |
| Bomb or arson attacks | |
| Violence against persons | incl. kidnapping, persons injured or killed |
| Clashes with police | |

For the assessment of the relevance of a document or sentence, it is sufficient that a reference to a protest event corresponds to one of the actions listed.

***PLEASE NOTE:***

A rule of thumb is that political protest is 'bottom up'. Actions which are conducted as part of usual institutional processes are thus not considered in this analysis. This refers, for example, to police operations, court cases or all kinds of legislative procedures. Also, protests by political authorities of whatever kind do not count as protest events.

With the exception of bomb threats, we exclude threats or planned events that are announced without any specified date. If the date is specified, we include them.

We don't evaluate whether a protest is legal or not. Every political protest, which meets an entry in the list of action forms above, is relevant.

## 2. IDENTIFICATION OF EVENTS

How do we identify protest events?

A relevant document may contain references to one protest event only, or it may contain references to more than one protest event. Moreover, the same protest event may be referred to several times at different places in the document. The events in the document can be distinguished on the basis of the following 'form-time-locality' rules:

1. Protests of different action forms define different events, e.g. the combination of a general strike with a demonstration in the capital city on the same day would define two separate protest events.
2. Protests of a given action form define different events, if they take place at different times, i.e. different days or clearly distinct moments on the same day (e.g. a series of demonstrations by the same actor, but on successive days) and/or in different locations (e.g. parallel demonstrations for the same goal that take place on the same day, but in different cities, correspond to as many protest events as cities are mentioned in the document). Different events of the same form (e.g. strike) taking place on the same day at different places in the same city count as one event, even if they are mentioned separately (e.g. strikes by different actors, such as metro drivers and airport controllers).

***PLEASE NOTE:***

To identify individual protest events, we apply the same list of action forms that we also use for the evaluation of the relevance (see action forms).

A particular problem is posed by what we call “campaign references”, which are general, summarizing references to several protest events. An example is the following sentence: “Today, teachers in three different parts of Spain protested.” According to our rules to separate protest events, we would expect to annotate several events, i.e. one for each of the “three different parts of Spain”, but there is no explicit information on these locations. There are two alternatives how we have to deal with this problem:

1. If the story presents explicit information on all the single events, e.g. “in Madrid, 10’000 teachers gathered in front of the parliament, while about 5000 took to the streets in Valencia and Barcelona”“, we annotate the three separate events, one each in Madrid, Barcelona and Valencia.
2. If it is clear that the ‘campaign reference’ does mention much more than the references to the single events, or if there is no further reference specifying the single events at all, we take this campaign reference as a separate protest event. In the example, we would annotate a protest event in Spain on the basis of the ‘campaign reference’ if less than three events are more detailed described.

Enumerations of protest events, e.g. ‘in 80 cities people took to the streets’ or this is the ‘5th bomb attack this year’, are considered as single protest events, if no further information on locations and dates is given.

An event which does qualify as a protest event on the basis of the form-time-locality rules, is not coded as a protest event. Also, it is no protest event if the actor performing the event is an ‘elite’-actor, who performs the action…

… as part of his daily routine (e.g. an interest association protesting (in a press conference, by writing a letter or by lobbying a decision-maker)

… exceptionally, but without mobilizing non-elite individuals (e.g. French mayors protesting against the law on gay marriages and refusing to implement it), or

… in inter-state relations (e.g. cyber-attacks by Russia against the Baltic states).

You do not have to annotate previously annotated events. However, if a new reference to a previously annotated event contains additional information, you should add this new information to the annotation of this event.

## 3. INDICATORS

The following table shows the list of indicators that have to be annotated per protest event as well as their descriptions.

| Label | Description |
|---|---|
| Event index | Index indicating separate events |
| Date of event | Date of event as specified in the document. |
| Location of event | Geographical reference associated with the protest event. |
| N. of participants | Information on the size of protest event in terms of participants. |
| Action form | Action form applied during a protest event. |
| Goal of protest | Political goals associated with the protest event. |
| Actors | Political groups organizing or participating in a protest event. |

***PLEASE NOTE:***

The event index refers to the number of the event as identified in step 2 (identification of events).

## 4. MANUAL

### Annotation instructions

### General Remarks

Most of the events should be quite straightforward to annotate. However, there are certainly also difficult decisions during the annotations, for which the following guidelines might be helpful.

1. *Overinterpretation:* We try to objectivize the information we read in the text as well as possible. This means that we only annotate what is explicitly mentioned and abstain from adding information from our common knowledge or interpretations of the information. For example, if we know that thousands of persons take part in a general strike (because, obviously, a general strike involves the entire economy of a country), but we do not explicitly read about this in the story, we do not enter this information on the number of participants. In a similar vein, if it is written in a story that a demonstration in a city is 'similar' in terms of the number of participants than another demonstration in another city, we do not enter this information if there is no other, explicit description of the number of participants. However, please, note that we use our own knowledge about detailed locations in order to code the name of cities (e.g. if 'Acropolis' is mentioned, but no city name, we code 'Athens' for the city, if we know that the Acropolis is located in Athens; or if 'Cotroceri Palace' is mentioned and we know that this is in Bukarest, we code 'Bukarest').
2. *Missing information:* To define a protest event, only an action form, a location (which can also be an entire country) and a time (which can also be an extended time period) need to be explicitly mentioned. All other indicators are 'optional' in the sense that we can leave them empty if no information is provided for them.
3. *Ambivalent information:* If there is ambivalent information, we rely on the description of protests by the journalist. For example, if a story reports about ethnic violence, but a

police report that is mentioned suggests that it is a criminal act (so no relevant protest), it depends on whether other information in the story explicitly links the violence to ethnic conflicts or not. If yes, the event nevertheless is a relevant protest.

During the annotation, each document has to be read twice:

In the first reading, all references to relevant protest events have to be identified. There are pre-annotated sentences, which are highlighted in yellow were found on the basis of relevant language patterns by an automated recognition, but you also have to search for sentences that are not pre-annotated but may still be relevant. Moreover, words describing locations are highlighted in green, words referring to times are highlighted in blue, and names of persons and organizations are highlighted in red.

***PLEASE NOTE:***

The first reading also serves the purpose to identify the different references, which refer to the same protest event. So please think about which references belong to the same event while doing this first reading.

Every event automatically receives an index, i.e. a number that allows us to separate the different events. During the annotation, you have to enter all variables at the level of events. Hence, you have to consider the information from all references on the same event for the coding of the indicators.

In the example shown in the following table, the first and second references do not refer to the same event, since there is a clear difference in time between the two events mentioned ("last year"). These references have different event indices (values 1 and 2). The references three and four, however, belong to the same event, which is indicated with the same event index of value 3.

| **References to a protest event** | **Event index** |
|---|---|
| "On Monday, several thousand students protested against the planned raise in tuition fees." | 1 |
| "It's the largest student protests since the demonstrations last year." | 2 |
| "Truckers blocked the main entrance to the city." | 3 |
| "50 trucks brought traffic on the highway to a standstill." | 3 |

During the second reading, the labels of the six indicators date, location, number of participants, action form, issue and actors have to be annotated for every protest event.

If a document contains no relevant protest, you have to click the `irrelevant story` checkbox before going to the next story.

**Dates**

If possible, we always indicate the full date in the format yyyy-mm-dd, i.e. the day, month and year separated by a point. This means that also indirect information on the exact date (e.g. "yesterday") have to be translated into an actual date with help of the date of publication. Hence, for the example of “yesterday”, if the date of publication is 02-09-2006, the date entry is 01-09-2006.

***PLEASE NOTE:***

If the date is not clear from the text, we enter the date of publication of the news story.

If only a longer time period is indicated (e.g. “the sit-in from last Wednesday on”, “the bombings in August 2003” or “the riots in 2003”), there are three ways to enter the date, depending on the information provided:

1. If a starting day is apparent as in the first example, we take this as the date of the protest event.
2. If no starting day is given, but the time period is equal or less than a week, we take the Monday of the corresponding week as date of the protest.
3. If no starting day is evident and the time period is longer than a week, we have to enter the time as is into the entry field named ‘Non-standard date’, e.g. ‘2003-08’ or ‘2003’ for the second and third example above. If possible, convert the dates you enter in the ‘Non-standard date’ field as number with the same format as the date in the date entry field.

**Locations**

We enter all different geographic names related to a protest event into the first text field. If possible, please indicate only the name of the city where the protest takes place. E.g. if a story reports on a ‘bomb attack at Belfast airport’, enter ‘Belfast’ and not ‘Belfast airport’ into the location entry field. In the second list selection field, you additionally have to select the country.

**Number of participants**

Here, all information on the size of a protest needs to be entered. This information can be present in the documents as numbers (e.g. ‘50 workers’) or words (e.g. ‘one thousand protesters’). Also, the information may be precise or only vague (e.g. ‘several demonstrators’ or ‘thousands of teachers’). Please use the following rules to enter the number of participants:

1. If the number is precise, please enter it in digits. Thus, ‘thousand’ would have to be entered as ‘1’000’.
2. If the information is vague, please use the conversion indicated in the following table.

| example of reference | number to enter |
|---|---|
| several, some, a few, a group of, a couple of | 5 |
| dozens, a number of | 50 |
| hundreds | 500 |
| thousands | 5’000 |
| tens of thousands | 50’000 |
| hundreds of thousands | 500’000 |
| millions | 5’000’000 |

***PLEASE NOTE:***

Only persons which are actively involved in the protests are considered participants. Thus, neither the police, spectators nor victims are considered as participants.

If there is more than one piece of information given, take the one that is more precise (e.g. if ‘hundreds’ and ‘600’ are mentioned in relation to the same protest event, enter ‘600’). If there are multiple pieces of information at the same level of precision (e.g. one number is published by the police and a different number by the organizers), please take the average.

We enter unclear indications such as ‘more than 3000’ or ‘over 20’000’ defensively. The examples thus are entered as 3’000 and 20’000, respectively. If there are indications on the size of protests you cannot locate in the conversion table such as ‘many’ or ‘a large crowd’, please leave the entry field ‘number of protesters’ empty and copy-paste the corresponding string plus a remark into the comment field.

**Action forms**

One of the following labels has to be selected for the action form indicator.

| **Label** | **Description** |
|---|---|
| Strikes | industrial action of any kind (incl. work stoppages, pickets) |
| Demonstrations | Demonstrations, marches, rallies, meetings, vigils and other non-confrontational gatherings |
| Petitions and related activities | petitions, letters, consumer activism, boycotts, symbolic protests (performances etc.) |
| Blockades and related activities | blockades, occupations, sit-ins, camps and other confrontational strategies that are related to a specific place |
| Violent protests | Sabotage, riots, destruction of private or public buildings, bomb or arson attacks, violence against persons, clashes with police, cyber-attacks |
| Other protests | |

**Issues**

For the annotation of the issues, one of the following labels has to be selected:

| **Label** | **Description** |
|---|---|
| Economics (private) | Economic claims addressed to firms/ employers: dismissal of staff, closure of firm/branch, labor conflict related to pay rise, pay cut etc. |
| Economics (public) | Economic claims addressed to public institutions, e.g. welfare, budget policies, agricultural policies, labor regulation, state regulation of the economy more generally |

| Label | Description |
|---|---|
| Environment | Environmental protection, Nuclear energy, Other forms of energy production, Infrastructure projects (e.g., transport), Animal rights |
| Cultural liberalism | Peace (questions of war & peace, nuclear and other conventional weapons, military infrastructure, military spending, military service), Women's rights (incl. equal treatment, abortion), LGTB (rights and recognition of lesbians, gay, transsexuals, bisexuals), International solidarity (development aid; anti-imperialism), Anti-racism, rights of migrants more generally, Squatters mobilization (for autonomous living and cultural spaces) |
| Regionalism | Separatism, regional independence, protection of regional interests, anti-regionalist counter-mobilization |
| Cultural conservatism | counter-mobilisation to all aspects of new social movement except for anti-racism and migrants' rights (which is in xenophobia) |
| Xenophobia | Right-wing extremism, racist mobilization (against foreigners or ethnic minorities), anti-Immigration |
| Political | Representation, corruption, electoral reform, institutional issues in general |
| Others | All other issues not covered by the previous categories |

***PLEASE NOTE:***

There can be none, one or multiple issues per event. In case of multiple issues, you should select all categories that apply.

**Actors**

We code all actors that are reported to call for, take part in or organize a given protest event. If you find such an actor, one of the following labels has to be selected:

| Label | Description |
|---|---|
| Parties (left) | Political parties from the left (e.g. Social Democrats, Greens or Communists) |
| Parties (right) | Political parties from the right (e.g. Liberals, Christian Democrats, Conservatives, Nationalists or Agrarians) |
| Parties (unknown) | Political parties for which the ideological orientation is not clear |
| Unions (private) | Private sectors unions |
| Unions (public) | Public sector unions |
| Unions (both) | Unions for both the private and public sector |
| Unions (unknown) | Unions for which sector affiliation is not clear |
| Other Organizations | Residual category for all other types of organized actors such as NGO's |
| Social groups (occupational) | Workers, teachers, lawyers, journalists, nurses |

| Label | Description |
| --- | --- |
| Social groups (students) | |
| Social groups (pensioners) | |

The categories contain two different types of actors. Categories 1 to 7 cover formal organizations, whereas the other categories cover (unorganized) social groups. In the case of organizations, we distinguish three types: political parties, trade unions, and all other types of organizations. Please refer to the lists in the Appendix for an overview over parties and unions.

***PLEASE NOTE:***

There can be none, one or multiple actors participating in an event. In case of multiple actors, you should select all categories that apply.

Political parties are divided into left-wing and right-wing parties. In case of doubt about this classification, use category 'political party (unknown)'. The decision to code parties as left-wing or right-wing can be based on explicit references in the text or the list of political parties provided in the appendix, see list of parties.

Note that the references in the text might explicitly refer to the political orientation of the parties (left-wing or right-wing) or to party families. We categorize party families as follows: 1. Left-wing = Social Democrats, Greens, New Left, Communists, Radical left 2. Right-wing = Conservatives, Christian Democrats, Liberals, Radical right

Unions are divided into public, private and both sectors unions. In case of doubt, use category 'Unions (unknown)'. Again, the decision to code unions as public or private can be based on explicit references in the text or on the list of unions in the appendix, see list of unions.

Note that references in the text might usually refer to the occupational groups organized by the union(s). Public sector covers all groups employed by the state, such as civil servants, police officers, teachers, doctors, nurses or fire fighters. Private sector covers all occupational groups employed by private businesses, such as industrial workers, service sector employees, miners, taxi drivers, etc.

Social groups are not organized groups of persons, which can be distinguished by a relevant societal characteristic. In these annotations, we only focus on the labor market position occupation, i.e. groups of people which are characterized by a job or by unemployment (='Social groups (occupational)'), as well as on pensioners (='Social groups (pensioners)') and students (='Social groups (students)').

Single persons and very general descriptions (e.g. ‘people’, ‘persons’ or ‘protesters’) are not considered as social groups.

If no specific organization is mentioned, we annotate actors as social groups. Hence, ‘doctors’ is a social group. If an ‘Association of doctors’ is mentioned, we annotate this actor as an organization.

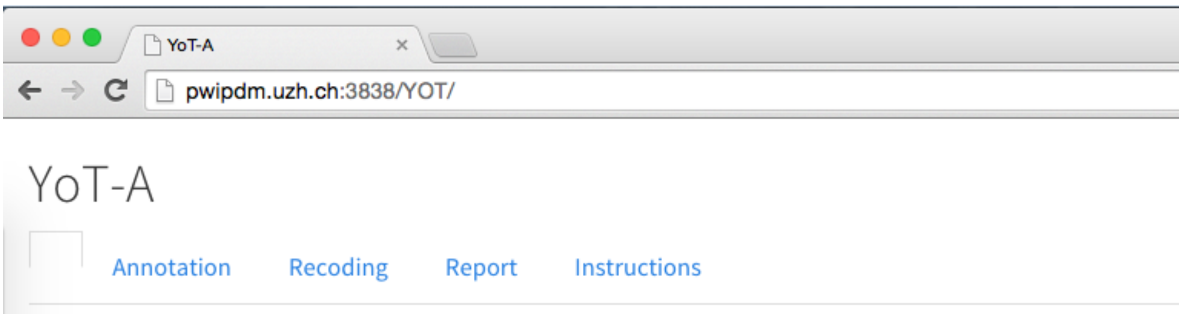
**Annotation interface**

The annotation is done in an online interface called “YoT-A”. It can be reached at <link to R-shiny app>. You can work with any browser.

***PLEASE NOTE:***

If you are annotating from outside the UZH network, you need to establish a VPN tunnnel first.

On the webpage, you’ll see a list of links, one of which should be your last name. Click on it to login. Upon login, you’ll see the entry page shown below.

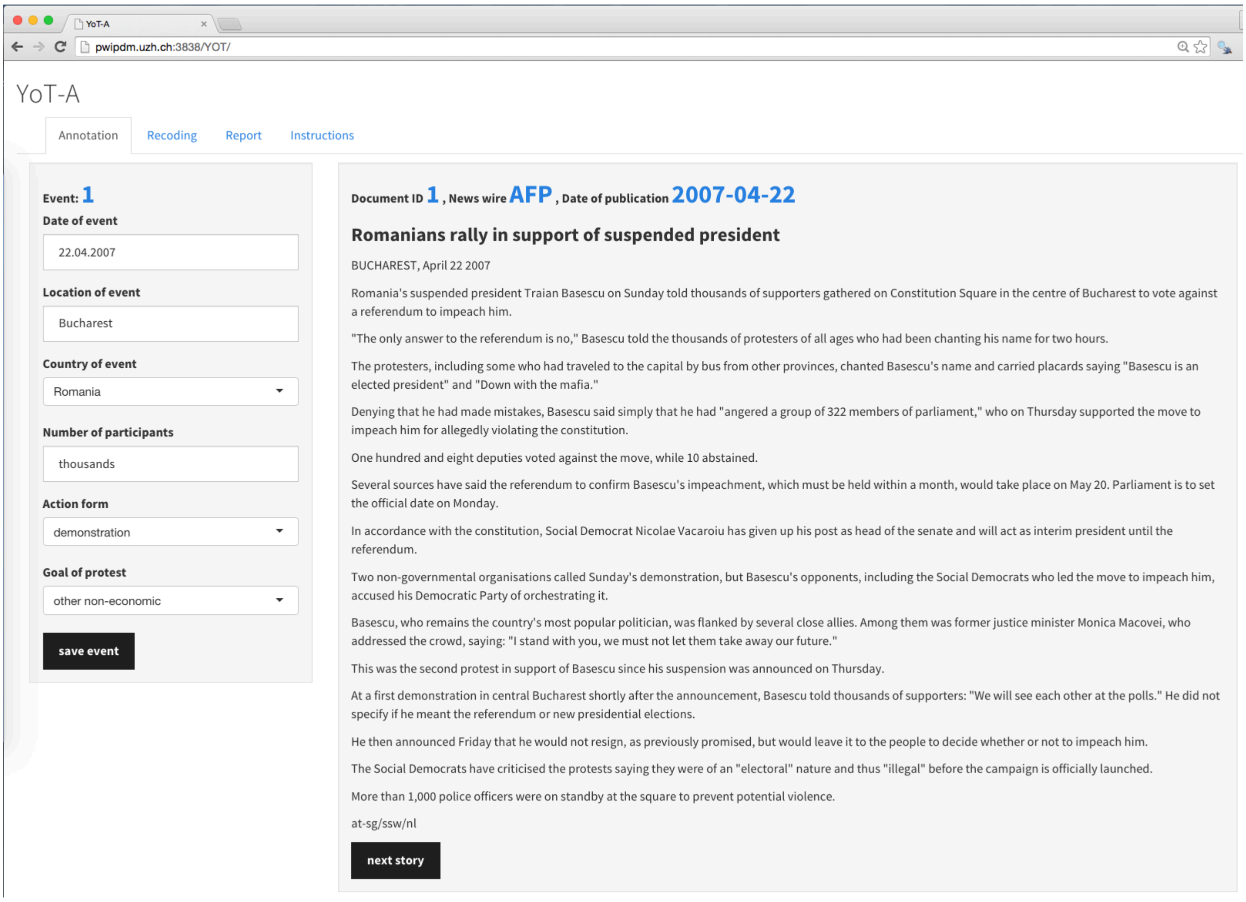



There are four panels:

1. Annotation: Annotate protest events here.
2. Recoding: Make corrections on and get an overview over your previous annotations.
3. Report: Fill out your weekly report.
4. Instructions: This file.

**Annotation**

In the annotation panel, you have the entry fields for the single indicators on the left and the news story (incl. document ID, news wire and date of publication) on the right (see image below). When starting an annotation session, you have to get a first report by clicking ‘next story’. Then, you should enter the information on all indicators for the first event on the right. If you are done, click ‘save event’ to save your annotation on this event and open a new form for a potential further event. If there is no further event, click on ‘next story’ to finish this document and start with a new one.




***PLEASE NOTE:***

Don’t forget to click ‘save event’, also if there is only one event per story. Otherwise, all your annotations for this event will be lost!

You can move around with the tab and arrow keys or with your mouse. The ‘location of event’ (text entry) and ‘number of participants’ (numeric entry) are simple entry fields. The ‘country of event’, ‘action form’ and ‘goal of protest’ are select lists where you have pre-defined options and autocompletion. The date entry field, finally, has a calendar function which pops up when you click on it.

**Recoding**

In the recoding panel, all your previous annotations are shown in a table. Very similar to an excel table, you can change the content of every cell. The table also offers autocompletion and you can delete whole rows (i.e. events) by a right mouse-click and selecting ‘remove row’ from the menu while pointing the cursor on the corresponding row.

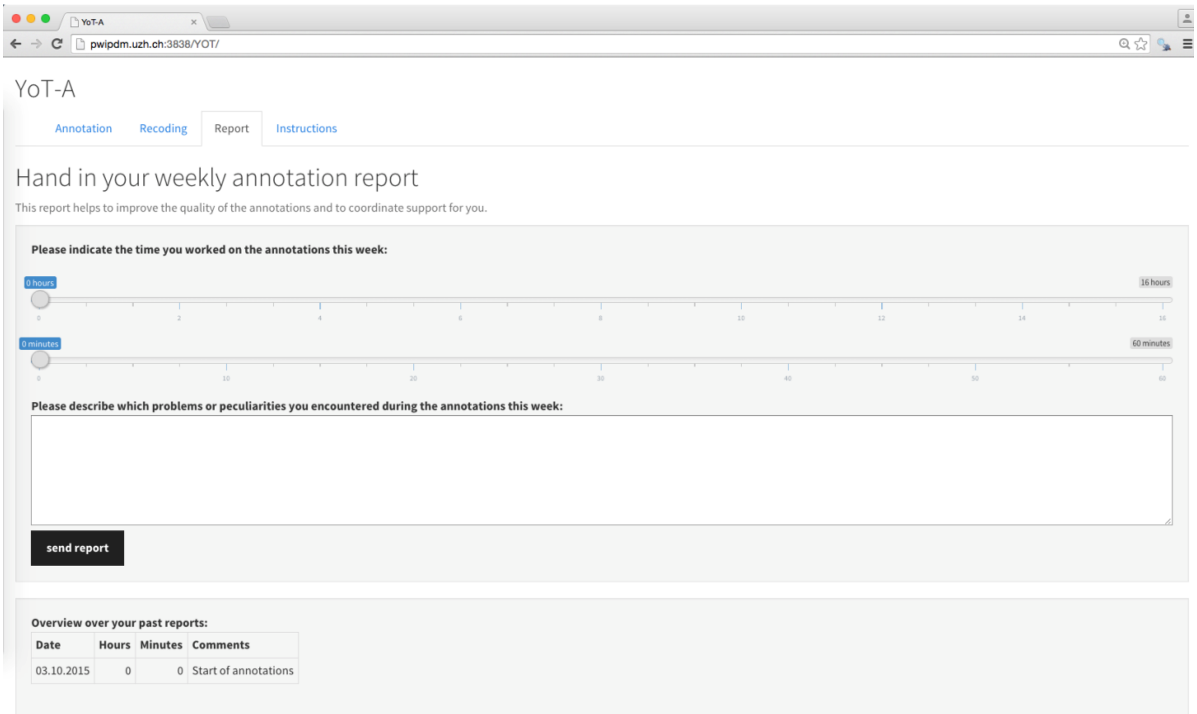
***PLEASE NOTE:***

If your most recent event annotations are not shown, you have to reload the page so the table is updated.

You can delete all rows except the first one. If you try to delete the first row, the table is not displayed properly (the message “Error: invalid ‘row.names’ length”) is shown. No data get lost, but you have to reload the webpage (not just the table!) to start working in the recoding table again.

### Report

In the report panel, you can hand in your weekly report (see image below). Please indicate the hours and minutes you worked since the last report and add comments on problems or peculiarities if necessary. At the bottom of the page, your past reports are shown for your reference.




***PLEASE NOTE:***

You have to fill out one report per week! If you forget to fill out a report, a reminder will be posted on the entry page.

## 5. APPENDIX

### List of parties

| Country | Name | Orientation |
|---|---|---|
| Austria | GRÜNE The Greens | Left |
| | KPÖ Austrian Communist Party | Left |

| Country | Name | Orientation |
|---|---|---|
| | SPÖ Austrian Social Democratic Party | Left |
| | BZÖ Alliance for the Future of Austria | Right |
| | FPÖ Austrian Freedom Party | Right |
| | LIF Liberal Forum | Right |
| | NEOS The New Austria | Right |
| | ÖVP Austrian People’s Party | Right |
| | TeamSTronach | Right |
| Belgium | AGALEV Live Differently | Left |
| | ECOLO Ecologists | Left |
| | groen! Green! | Left |
| | PS Francophone Socialist Party | Left |
| | SP Flemish Socialist Party | Left |
| | sp.a Socialist Party Different | Left |
| | sp.a-SPIRIT Socialist Party Different – Spirit | Left |
| | SPIRIT Social, Progressive, International, Regionalist, Integrally Democartic and Forward-Looking | Left |
| | CD&V Christian Democratic and Flemish | Right |
| | CVP Christian People’s Party | Right |
| | LDD List Dedecker | Right |
| | MR Reform Movement | Right |
| | N-VA New Flemish Alliance | Right |
| | openVLD Open Flemish Liberals and Democrats | Right |
| | PRL-FDF Liberal Reformation Party - Francophone Democratic Front | Right |
| | PRL-FDF-MCC Liberal Reformation Party - Francophone Democratic Front - Citizens’ Movement for Change | Right |
| | PSC Christian Social Party | Right |
| | VB Flemish Bloc | Right |
| | VB Flemish Interest | Right |
| | VLD Flemish Liberals and Democrats | Right |
| | VU Peoples’ Union | Right |
| | VU-ID21 People’s Union - Complete Democracy for the 21st century | Right |
| Bulgaria | ABV Alternative for Bulgarian Revival | Left |
| | BSP Bulgarian Socialist Party | Left |
| | DL Democratic Left | Left |
| | DPS Movement for Rights and Freedom | Left |
| | KEL Euroleft Coalition | Left |
| | KzB Coalition for Bulgaria | Left |
| | PKE Political Club “Ekoglasnost” | Left |
| | ATAKA National Union Attack | Right |
| | BNS Bulgarian People’s Union | Right |
| | DBG Bulgaria Citizens Movement | Right |
| | DSB Democrats for a Strong Bulgaria | Right |
| | GERB Citizens for European Development of Bulgaria | Right |
| | NDSV National Movement Simeon the Second | Right |
| | NFSB National Front for the Salvation of Bulgaria | Right |
| | ODS United Democratic Forces | Right |
| | SDS Union of Democratic Forces | Right |

| Country | Name | Orientation |
|---|---|---|
| | SK Blue Coalition | Right |
| | VMRO-BND Bulgarian National Movement | Right |
| Cyprus | AKEL Progressive Party of the Working People | Left |
| | EDEK United Democratic Union of Cyprus | Left |
| | KISOS Social Democrats’ Movement | Left |
| | KOP Ecological and Environmental Movement | Left |
| | DIKO Democratic Party | Right |
| | DISY Democratic Coalition | Right |
| | KED Free Democrats Movement | Right |
| | KF Liberal Party | Right |
| Czech republic | CSSD Czech Social Democratic Party | Left |
| | KSCM Communist Party of Bohemia and Moravia | Left |
| | SZ Green Party | Left |
| | USVIT Dawn of Direct Democracy | Left |
| | KDU-CSL Christian and Democratic Union - Czech People’s Party | Right |
| | ODA Civic Democratic Alliance | Right |
| | ODS Civic Democratic Party | Right |
| | SPR-RSC Association for the Republic – Republican Party of Czechoslovakia | Right |
| | SVOBODNI Party of Free Citizens | Right |
| | TOP09 Tradition, Responsibility, Prosperity 09 | Right |
| | VV Public Affairs | Right |
| Denmark | CD Centre Democrats | Left |
| | EL Red-Green Unity List | Left |
| | FolkB People’s Movement Against the EU | Left |
| | SD Social Democratic Party | Left |
| | SF Socialist People’s Party | Left |
| | DF Danish People’s Party | Right |
| | KrF OR KF Christian People’s Party | Right |
| | LA Liberal Alliance | Right |
| | NY New Alliance | Right |
| | RV Radical Party | Right |
| | V Liberals | Right |
| Estonia | EER Estonian Greens | Left |
| | EK Estonian Center Party | Left |
| | M People’s Party Moderates | Left |
| | SDE Social Democratic Party | Left |
| | Electoral Union ‘Our Home is Estonia’ | Right |
| | EME Estonian Rural People’s Party | Right |
| | ER Estonian Reform Party | Right |
| | ERL Estonian People’s Union | Right |
| | EÜRP Estonian United People’s Party | Right |
| | EVE Estonian Free Party | Right |
| | IRL Pro Patria and Res Publica Union | Right |
| | K Estonian Center Party | Right |
| | KE Coalition Party | Right |
| | KMÜ Coalition Party and Rural Union | Right |

| Country | Name | Orientation |
|---|---|---|
| Finland | SDP Finnish Social Democrats | Left |
| | VAS Left Wing Alliance | Left |
| | VL Green Union | Left |
| | VIHR Green League | Left |
| | KD Christian Democrats in Finland | Right |
| | KESK Finish Center Part | Right |
| | KOK National Coalition Party | Right |
| | NSP Progressive Finnish Party, also known as Young Finns | Right |
| | PS True Finns | Right |
| | RKP/SFP Swedish People's Party | Right |
| | SK Finnish Centre | Right |
| | SKL Finnish Christian Union | Right |
| | SMP Finnish Rural Party | Right |
| France | GE Ecology Generation | Left |
| | ENS Ensemble - Together | Left |
| | Les Verts EELV The Greens | Left |
| | PCF French Communist Party | Left |
| | PG Left Party | Left |
| | PS Socialist Party | Left |
| | PRG Radical Party of the Left | Left |
| | DLR Debout La Republic - France Arise | Right |
| | FN National Front | Right |
| | UDF Union for French Democracy | Right |
| | UMP Union for Popular Movement | Right |
| | MPF Movement for France | Right |
| | MODEM Democratic Movement | Right |
| | NC New Center | Right |
| | PRV Radical Party | Right |
| Germany | 90/Greens Alliance'90/Greens | Left |
| | L-PDS The Left. Party of Democratic Socialism | Left |
| | LINKE The Left | Left |
| | PDS Party of Democratic Socialism | Left |
| | Pirate party | Left |
| | SPD Social Democratic Party of Germany | Left |
| | AfD Alternative for German | Right |
| | CDU/CSU OR UNION Christian Democratic Union/Christian Social Union | Right |
| | DVU German People's Union | Right |
| | FDP Free Democratic Party | Right |
| | NPD National Democratic Party | Right |
| | Republikaner Republicans | Right |
| Great britain | Greens Green Party | Left |
| | Labour Labour Party | Left |
| | SNP Scottish National Part | Left |
| | SF Ourselves Alone | Left |
| | CONS Conservatives | Right |
| | DUP Democratic Unionist Party | Right |
| | LibDems Liberal Democrats | Right |

| Country | Name | Orientation |
| --- | --- | --- |
| Greece | DIKKI Democratic Social Movement | Left |
| | DIMAR Democratic Left | Left |
| | KKE Communist Party of Greece | Left |
| | PASOK Panhellenic Socialist Movement | Left |
| | SYP Progressive Left Coalition | Left |
| | SYRIZA Coalition of the Left, Movements and Ecology | Left |
| | To Potami (the river) | Right |
| | ANEL – Independent Greeks | Right |
| | Golden Dawn | Right |
| | LAOS Popular Orthodox Rally | Right |
| | ND New Democracy | Right |
| | Pola Political Spring | Right |
| Hungary | DK Democratic Coalition | Left |
| | E14 Együtt 2014 - Together 2014 | Left |
| | Liberalisok | Left |
| | LMP Politics Can Be Different | Left |
| | MSzDP Hungarian Social Democratic Party | Left |
| | MSzP Hungarian Socialist Party | Left |
| | SzDSz Alliance of Free Democrats | Right |
| | Fidesz | Right |
| | FKgP Independent Smallholders' Party | Right |
| | Jobbik Movement for a Better Hungary | Right |
| | KDNP Christian Democratic People's Party | Right |
| | MDF Hungarian Democratic Forum | Right |
| | MIÉP Hungarian Justice and Life Party | Right |
| Iceland | A Social Democratic Party | Left |
| | Ab People's Alliance | Left |
| | S The Alliance | Left |
| | TfFf Awakening of the Nation | Left |
| | VGF Left Green Movement | Left |
| | F Progressive Party | Right |
| | FF Liberal Party | Right |
| Ireland | DLP Democratic Left Party | Left |
| | Greens Green Party | Left |
| | Labour Labour Party | Left |
| | PBPA People Before Profit Alliance | Left |
| | SP Socialist Party | Left |
| | Familiy of the Irish - Fine Gael | Right |
| | Fianna Fail | Right |
| | PD Progressive Democrats | Right |
| Italy | DS Democrats of the Left | Left |
| | FdV Green Federation | Left |
| | M5S Five Star Movement | Left |
| | Olive Tree | Left |
| | Pannella-Sgarbi List | Left |
| | PdCI Party of Italian Communists | Left |
| | PD Democratic Party | Left |

| Country | Name | Orientation |
|---|---|---|
| | PDS Democratic Party of the Left | Left |
| | PRC Communist Refoundation Party | Left |
| | RC Civil Revolution | Left |
| | RI Italian Renewal | Left |
| | RnP Rose in the Fist | Left |
| | SEL Left Ecology Freedom | Left |
| | The Girasole (“Sunflower”) | Left |
| | VdA Aosta Valley | Left |
| | AD Democratic Alliance | Right |
| | ALD Autonomy Liberty Democracy (Aosta Valley) | Right |
| | AN National Alliance | Right |
| | Autonomy Progress Federalism Aosta Valley | Right |
| | CCD Christian Democratic Centre | Right |
| | CD Democratic Centre | Right |
| | Fdl Fratelli d’Italia | Right |
| | FI Forza Italia | Right |
| | IdV List Di Pietro - Italy of Values | Right |
| | LN Northern League | Right |
| | M-DL Daisy Democracy is Freedom | Right |
| | NCD New Center-Right | Right |
| | PD Democratic Party | Right |
| | PPI Italian Popular Party | Right |
| | SC Civic Choice | Right |
| | SVP South Tyrolean People’s Party | Right |
| | UDC Union for Christian and Center Democrats | Right |
| | UdC Union of the Center | Right |
| | White Flower | Right |
| Latvia | LKS Latvian Russian Union | Left |
| | LSDA Latvian Social Democratic Alliance | Left |
| | LSP Latvian Socialist Party | Left |
| | LVP Latvian Unity Party | Left |
| | PCTVL For Human Rights in a United Latvia | Left |
| | SDPS Social democratic Party “Harmony” | Left |
| | TSP National Harmony Party | Left |
| | DPS Democratic Party ‘Saimnieks’ | Right |
| | JP New Party | Right |
| | LC Latvian Way Union | Right |
| | LPP Latvia’s First Party | Right |
| | LZS-LKDS-LDP Coalition of Latvian Farmers’ Union, Latvian Christian Democratic Union and Democratic Party of Latgale | Right |
| | NA LNNK National Alliance “all For Lativa” | Right |
| | NSL For Lativa from the Heart | Right |
| | TB-LNNK For Fatherland and Freedom - Latvian National Independence Movement | Right |
| | TKL-ZP Popular Movement for Latvia-Zigerista Party | Right |
| | TUB For Fatherland and Freedom | Right |
| | V Vienotiba Unity | Right |

| Country | Name | Orientation |
|---|---|---|
| | ZZS Greens' and Farmers' Union | Right |
| Lithuania | BSDK A. Brazauskas Social Democratic Coalition | Left |
| | LDDP OR DP Lithuanian Democratic Labour Party | Left |
| | LSDP Lithuanian Social Democratic Party | Left |
| | LVZS Lituanian Unon of Greens and Farmers | Left |
| | UdL Working for Lithuania | Left |
| | DK The Way of Courage | Right |
| | LCS Lithuanian Centre Union | Right |
| | LiCS Liberal and Centre Union | Right |
| | LKDP Lithuanian Christian Democratic Party | Right |
| | LLRA Election Action of Lithuania's Poles | Right |
| | LLS Lithuanian Liberal Union | Right |
| | LRLS Liberal Movement | Right |
| | LVP Lithuanian Peasants Party | Right |
| | LVŽS Lithuanian Peasant and Green Union | Right |
| | NS New Union (Social Liberals) | Right |
| | PTT OR TT Order and Justice | Right |
| | TS-LKD Homeland Union - Lithuanian Christian Democrats | Right |
| | UTT For Order and Justice | Right |
| | VNDS Union of Peasants and New Democracy Party | Right |
| Luxembourg | GLEI-GAP Green Left Ecological Initiative - Green Alternative | Left |
| | LSAP/POSL Socialist Workers' Party | Left |
| | The Greens | Left |
| | The Left | Left |
| | CSV/PCS Christian Social People's Party | Right |
| | DP/PD Democratic Party | Right |
| Malta | PL Labour Party | Left |
| | PN Nationalist Party | Right |
| Netherlands | D'66 Democrats'66 | Left |
| | GL Green Left | Left |
| | PvdA Labour Party | Left |
| | PvdD Party for the Animals | Left |
| | SP Socialist Party | Left |
| | CDA Christian Democratic Appeal | Right |
| | CU Christian Union | Right |
| | LN Livable Netherlands | Right |
| | LPF List Pim Fortuyn | Right |
| | PVV Party of Freedom | Right |
| | SGP Reformed Political Party | Right |
| | VVD People's Party for Freedom and Democracy | Right |
| Norway | DnA OR AP Norwegian Labour Party | Left |
| | SV OR V Socialist Left Party | Left |
| | MDG Green Party | Left |
| | FrP Progress Party | Right |
| | H Conservative Party | Right |
| | KrF Christian People's Party | Right |
| | Sp Centre Party | Right |

| Country | Name | Orientation |
| --- | --- | --- |
| | V Liberal Party | Right |
| Poland | SLD Democratic Left Alliance | Left |
| | SLD-UP Coalion of the Democratic Left Alliance and the Union of Labour | Left |
| | RP Your Movement - Palikots Movement | Left |
| | UP Union of Labour | Left |
| | AWS Electoral Action ‘Solidarity’ | Right |
| | LPR League of Polish Families | Right |
| | MN German Minority | Right |
| | PO Civic Platform | Right |
| | PSL Polish Peasants’ Party | Right |
| | KNP Congress of the New Right | Right |
| | PO Civic Platform | Right |
| | PiS Law and Justice | Right |
| | PR Poland Together | Right |
| | PSL Polish People’s Party | Right |
| | ROP Movement for the Reconstruction of Poland | Right |
| | SP United Poaland | Right |
| | SRP Self-Defence of the Polish Republic | Right |
| | UW Freedom Union | Right |
| Portugal | BE Left Bloc | Left |
| | CDU Unified Democratic Coalition | Left |
| | PCP Portuguese Communist Party | Left |
| | PEV Ecologist Party ‘The Greens’ | Left |
| | PS Socialist Party | Left |
| | CDS-PP Social Democratic Center-Popular Party | Right |
| | MPT Earth Party | Right |
| | PP Partido Popular | Right |
| | PSD Social Democratic Party | Right |
| Romania | PDSR OR PSD Party of Social Democracy of Romania | Left |
| | PP -DD Peoples Party - Dan Diaconescu | Left |
| | USD Social Democratic Union | Left |
| | UNPR National Union for the Progress of Romania | Left |
| | ADA Justice and Truth Alliance | Right |
| | ALDE Alliance of Liberals and Democrats | Right |
| | CDR Democratic Convention of Romania | Right |
| | PC Conservative Party | Right |
| | PD Democratic Party | Right |
| | PDL Democratic Liberal Party | Right |
| | PLR Liberal Reformist Party | Right |
| | PMP Peoples Movement Party | Right |
| | PNL National Liberal Party | Right |
| | PRM Greater Romania Party | Right |
| | UDMR Hungarian Democratic Union | Right |
| Slovakia | KSS Communist Party of Slovakia | Left |
| | SDL’ Party of the Democratic Left | Left |
| | Smer Direction-Social Democracy | Left |
| | ANO Alliance of the New Citizen | Right |

| Country | Name | Orientation |
|---|---|---|
| | HZDS Movement for a Democratic Slovakia | Right |
| | KDH Christian Democratic Movement | Right |
| | MH Bridge | Right |
| | NOVA New Majority | Right |
| | OLaNO Ordinary People and Independent Personalities | Right |
| | SaS Freedom and Solidarity | Right |
| | SDK Slovak Democratic Coalition | Right |
| | SDKÚ Slovak Democratic and Christian Union | Right |
| | SDKÚ-DS Slovak Democratic and Christian Union - Democartic Party | Right |
| | Siet Network | Right |
| | SNS Slovak National Party | Right |
| | SMK/MKP Party of the Hungarian Coalition | Right |
| | SOP Party of Civic Understanding | Right |
| Slovenia | DeSUS Pensioners Party | Left |
| | PS Zoran Jankovic's List - Positive Slovenija | Left |
| | SD Social Democratic Party | Left |
| | SDSS Social-Democratic Party of Slovenia | Left |
| | SMC Pary of Miro Cerar | Left |
| | ZaAB Alliance of Alenka Bratusek | Left |
| | ZL United Left | Left |
| | ZLSD Associated List of Social Democrats | Left |
| | SDS Democratic Party (NOT SOCIAL DEMOCRATS) | Right |
| | For Real | Right |
| | LDS Liberal Democracy of Slovenia | Right |
| | LGV Gregor Virant's Civic List | Right |
| | Nsi New Slovenian Christian People's Party | Right |
| | SKD Slovenian Christian Democrats | Right |
| | SLS Slovenian People's Party | Right |
| | SLS-SKD Slovenian People's Party | Right |
| | SNS Slovenian National Party | Right |
| Spain | Amaiur | Left |
| | BNG Galician Nationalist Bloc | Left |
| | EA Basque Solidarity | Left |
| | ERC Republican Left of Catalonia | Left |
| | ICV Iniatives for Catalonian Greens | Left |
| | PODEMOS | Left |
| | PSOE Spanish Socialist Workers' Party | Left |
| | IU United Left | Left |
| | CC Canarian Coalition | Right |
| | C's Ciudadanos - Citizens | Right |
| | CiU Convergence and Union | Right |
| | EAJ/NPV Basque Nationalist Party | Right |
| | PP Partido Popular | Right |
| | Union, Progress and Democracy (UPyD) | Right |
| Sweden | FI Feminist Initiative | Left |
| | MP Green Ecology Party | Left |
| | SAP Social Democratic Labour Party | Left |

| Country | Name | Orientation |
|---|---|---|
| | V Left Party | Left |
| | CP Centre Party | Right |
| | FP Liberal People’s Party | Right |
| | Kd Christian Democrats | Right |
| | M Moderate Party | Right |
| | SD Sweden Democrats | Right |
| Switzerland | GPS/PES Green Party of Switzerland | Left |
| | PdAS/PdTS Swiss Labour Party | Left |
| | SPS/PSS Social Democratic Party of Switzerland | Left |
| | BDP Conservative Democratic Party | Right |
| | CSP/PCS Christian Social Party | Right |
| | CVP/PDC Christian Democratic People’s Party of Switzerland | Right |
| | EDU/UDF Federal Democratic Union | Right |
| | EVP/PEV Protestant People’s Party of Switzerland | Right |
| | FDP/PRD Radical Democratic Party | Right |
| | GLP Green Liberal Party | Right |
| | LdT Ticino League | Right |
| | LdU/AdI Independents’ Alliance | Right |
| | LPS/PLS Liberal Party of Switzerland | Right |
| | SD/DS Swiss Democrats | Right |
| | SVP/UDC Swiss People’s Party | Right |

## List of trade unions

| Country | Abbreviation | Name, description | Type |
|---|---|---|---|
| Austria | GdG-KMSfB | Union of Municipal Employees, Art, Media, Sport and Freelance Workers | both |
| | OeGB (ÖGB) | Austrian Federation of Trade Unions | both |
| | GBH | Union of Construction and Woodworkers | private |
| | GPA-DJP | Union of Salaried Private Sector Employees and Printers, Journalists and Paper Worker | private |
| | GPF | Union of Postal and Telecommunications Workers | private |
| | PRO-GE | Union of Production Workers | private |
| | vida | Transport and Service Union | private |
| | GÖD | Union of Public Services | public |
| Belgium | CSC | Confederation of Christian Trade Unions | both |
| | CSC | Transcom (Transports and Communication) | both |
| | FGBT | General Federation of Belgian Labour | both |
| | ACV-CSC-Metea | Metal and Textile | private |
| | CSC | Bâtiments industrie et Energie | private |
| | CSC | Alimentations et services | private |
| | CSC - CNE | Centrale Nationale des Employés | private |
| | CSC | Services Publics | public |
| | CSC | Enseignement | public |
| | ACLVB/CGSLB | General Confederation of Liberal Trade Unions of Belgium | both |

| Country | Abbreviation | Name, description | Type |
| --- | --- | --- | --- |
| | FGBT - UBT | Transport | both |
| | FGBT - CG | Centrale Générale | private |
| | FGBT - Horval | Alimentations - Horeca - Services | private |
| | FGBT - MWB | Métal | private |
| | FGBT - SETCa | Employés, Techniciens et Cadres | private |
| | FGBT - CGSP | Services Publics | public |
| Bulgaria | CITUB/KNSB | Confederation of Independent Trade Unions in Bulgaria | both |
| | CL Podkrepa | Confederation of Labour | both |
| Cyprus | DEOK | Cyprus Democratic Labour Federation | both |
| | PEO | Pancyprian Federation of Labour | both |
| | SEK | Cyprus Workers' Confederation | both |
| | ETYK | Cyprus Bank Employees Union | private |
| | SEK | Cyprus Workers Confederation | private |
| | OELMEK | Cyprus Secondary Education Greek Teachers Association | public |
| | OLTEK | Cyprus Vocational Education Greek Teachers Association | public |
| | PASYDY | Pancyprian Public Sector Workers Federation | public |
| | POED | Pancyprian Greek Teachers Association | public |
| Czech Republic | ASO | Association of Independent Trade Unions | both |
| | CMKOS | Czech-Moravian Confederation of Trade Unions | both |
| | KUK | Confederation of Art and Culture | both |
| | Odborové | Railway Workers' Union | both |
| | SOSaD, | Union of Services and Transport) | both |
| | CČR | Trade Union of Workers in the Chemical Industry | private |
| | KOVO | Trade unon for steel, engineering and electronics | private |
| | OSCPSPT | Trade Union of Workers in the Chemical, Paper, Glass and Print Industries | private |
| | OSK | Trade Union of Workers in the Metal Industry | private |
| | OSPZV | Trade Union of Workers in Agriculture and Food | private |
| | LOK | Medical Doctors Trade Union Club | public |
| | OSZSPČR | Trade Union of the Health Service and Social Care | public |
| Denmark | HK | The Union of Commercial and Clerical Employees in Denmark | both |
| | AC | Danish Confederation of Professional Associations | both |
| | FOA | Trade and Labour | both |
| | LO | Danish Federation of Trade Unions | both |
| | DANSK METAL | The Danish Metal Wrokers' Union | private |
| | Finansforbundet | The Financial Services Union | private |
| | 3F | The United Federation of Danish Workers | public |
| | BUPL | The Danish National Federation of Early Childhood Teachers and Youth Educators | public |
| | Danmarks Laererforening | The Danish Union of Teachers | public |
| | Dansk Socialradgiverforening | Danish Association of Social Workers | public |
| | Dansk Sygeplejerad | The Danish Nurses Organization | public |
| | FTF | The Confederation of Salaried Employees' and Civil Servants' Organisations | public |
| Estonia | EAKL | The Estonian Trade Union Confederation | both |

| Country | Abbreviation | Name, description | Type |
|---|---|---|---|
| | EAL | journalists' union | both |
| | EEAÜL | energy union | both |
| | ETTA | Road transport workers | both |
| | RTAL | broadcasting union | both |
| | TALO | Estonian Employees' Unions' Confederation | both |
| | EKTAL | light industries unon | private |
| | EMTAL | metal workers' union | private |
| | ROTAL | state and local governnment employees | public |
| | TTAÜ | customs officials union | public |
| Finland | TeMe | Theatre and Media Employees in Finland TeMe | boh |
| | AKAVA | Graduate-level employee unions | both |
| | AKT | Finnish Transport Workers' Union AKT | both |
| | SAK | Central Organisation of Finnish Trade Unions | both |
| | YLL | General Union of Journalists YLL | both |
| | STTK | Confederation of Unions for Academic Professionals in Finland | both |
| | IAU | Finnish Aviation Union IAU | private |
| | PAM | Service Union United PAM | private |
| | SEL | Finnish Food Workers' Union SEL | private |
| | SLSY | Finnish Cabin Crew Union SLSY | private |
| | SM-U | Finnish Seamen's Union SM-U | private |
| | TEAM | Industrial Union TEAM | private |
| | TEK | engineers | private |
| | Union | Finnish Electrical Workers' Union | private |
| | JHL | Trade Union for the Public and Welfare Sectors JHL | public |
| | OAJ | Teachers union | public |
| | TEHY | Healthc sector union | public |
| France | CFDT | Confédération Française Démocratique du Travail | both |
| | CFDT | Union syndicale des journalistes | both |
| | CFE-CGC | Confédération Française de l'Encadrement - Confédération Générale des Cadres | both |
| | CFTC | Confédération Française des Travailleurs Chrétiens | both |
| | CGT | Confédération générale du travail | both |
| | CNT | Confédération nationale du travail | both |
| | FO | Force Ouvrière | both |
| | SUD | Solidaires Unitaires Démocratiques | both |
| | UNSA | Union nationale des syndicats autonomes | both |
| | SIPN | Alliance Police nationale | public |
| | SM | Syndicat de la Magistrature | public |
| | USM | Union syndicale des magistrats | public |
| Germany | DGB | German Federation of Trade Unions | both |
| | EVG Eisenbahn- und Verkehrsgewerkschaft | Railway and Transport Services | both |
| | Verdi | United Services Union | both |
| | EVG | Railway and transport union | private |
| | Gewerkschaft Nahrung-Genuss-Gaststätten | Union Food, Beverage and Catering | private |
| | IG Bauen-Agrar-Umwelt | Union Construction, Agriculture, Environment | private |

| Country | Abbreviation | Name, description | Type |
|---|---|---|---|
| | IG Bergbau, Chemie, Energie | Union Mining, Chemical, Energy | private |
| | IGMETALL | Industrial Union of Metalworkers’ | private |
| | Gewerkschaft der Polizei | Police union | public |
| | Gewerkschaft Erziehung und Wissenschaft | Union Education and Science | public |
| Greece | GENOP | Electricity Sector | both |
| | PAME | All-Workers Militant Front | both |
| | CSEE | Greek Confederation of Labour | private |
| | OIPAE | Drivers | private |
| | OTOE | Banking Sector | private |
| | ADEDY | Civil Servants’ Confederation | public |
| Hungary | ASZSZ | Autonomous Trade Union Confederation | both |
| | LIGA | Confederation of Hungarian Trade Unions | both |
| | MOSZ | Confederation of Hungarian Trade Unions | both |
| | MSZOSZ | Confederation of Hungarian Trade Unions | private |
| | ÉSZT | Public Service (Health, Teachers …) | public |
| | SZEF | Higher Education and Research | public |
| Iceland | ASI | Confederation of Labour | both |
| | BSRB | Confederation of State and Municipal Employees of Iceland | public |
| Ireland | ICTU | Irish Congress of Trade Unions | both |
| | INMO | Irish Nurses and Midwives Organization | both |
| | Unite | (UK-based union) | both |
| | Mandate | Retail workers’ union | private |
| | TEEU | Irish Technical Engineering and Electrical Union | private |
| | Impact | Public sector union | public |
| Italy | FILT | Transport | both |
| | FeNEAL | Constuction and wood | private |
| | FILCAMS | Retail, restaurant, and hotel sectors | private |
| | FILCEM | Chemical, energy and manufacturing | private |
| | FILLEA | Construction | private |
| | FILTEA | Textile industry | private |
| | FIOM | Metalworkers | private |
| | FLAI | Farmworkers | private |
| | SLC | Communications | private |
| | FLC | Education | public |
| | FP | Public sector | public |
| | SPI | Pensioners | public |
| | CGIL | Italian General Confederation of Labour | both |
| | CISL | Confederation of Trade Unions in Italy | both |
| | UIL | Italian Workers Union | both |
| | UILT | Transports | both |
| | UIL COM | Communications | private |
| | UILA | Food and Agriculture | private |
| | UILCA | Bank, insurances, and tax collectors | private |
| | UILM | Metalworkers | private |
| | UIL Post | Postal workers | public |

| Country | Abbreviation | Name, description | Type |
|---|---|---|---|
| | UIL Scuola | Education | public |
| Latvia | LADAF | Latvian Aviation Workers' Federation | both |
| | LBAS | Trade Union Confederation | both |
| | LBAS | Latvijas Brivo arodbiedribu savieniba | both |
| | LCA | Latvian Builders' Trade Union | both |
| | LCDAA | Latvian Road workers trade association | both |
| | LDzSA, | Railway and transport workers union | both |
| | LIZDA | Latvian Education and Science Workers Union | both |
| | LKDAF | Latvian Culture Workers' Federation | both |
| | LLPNA | Latvian agricultural and food industry trade association | both |
| | LVPUFDA | Latvian State institutions, local governments, businesses and financial Workers Union | both |
| | LAB Energija | Latvian trade union "ENERGY" | dont know |
| | LIZDA | Education and Science Sector | dont know |
| | LMNA | Latvian Forest Industries Association trade | dont know |
| | LTFJA | Latvian Merchant Navy seamen union | dont know |
| | LIA | Latvian Trade Union of Industrial Sector | private |
| | LTDA | Latvian Commercial Workers Union | private |
| | LAKRS | Public services and transport employees | public |
| | LAPA | Latvian United Police Trade Union | public |
| | LPDA | Latvian Local Government Workers Union | public |
| | LSAB | Latvian Communications Workers Union | public |
| | LVSADA | Health and social work sector | public |
| | UTAF | Latvian Water Transport Federation of Trade Unions | private |
| Lithuania | LBMADPS | Lithuanian furniture and wood processing workers in the trade union | both |
| | LDF | Lithuanian Labour Federation | both |
| | lgpf | Lithuanian railway trade union federation | both |
| | litmetal | Lithuanian Metalworkers Trade Unions | both |
| | LKADPSF | Lithuanian road and transport workers' trade union federation | both |
| | LKDPF | Lithuanian Culture Workers Trade Unions Federation | both |
| | LKKDPS | Lithuania commerce and cooperation of workers' trade union | both |
| | LLPPS | Lithuanian Light Industry Trade Union | both |
| | LMMPDPSF | Lithuanian Forest and Wood Workers Trade Unions Federation | both |
| | lpsdps | Lithuanian service workers trade union | both |
| | LPSK | Lithuanian Trade Union Confederation | both |
| | lrdps | Lithuania communications workers' union | both |
| | LRTKDPS | Lithuanian radio and television creative staff professional association | both |
| | lvtdpsf | Lithuanian Water Transport Workers Federation of Trade Unions | both |
| | lzudps | Lithuanian agricultural workers trade union federation | both |
| | oprofsajunga | Lithuanian National Opera and Ballet Theatre, the trade union | both |
| | relektronika | Lithuanian radio electronics industry trade union federation of | both |

| Country | Abbreviation | Name, description | Type |
|---|---|---|---|
| | | organizations | |
| | Solidarumas | Lithuanian Labour Center | both |
| | LMPS | Lithuanian Trade Union of Food Producers | private |
| | lzs | Lithuanian Union of Journalists | private |
| | pramprof | Lithuanian industrial trade union federation | private |
| | LMPS | Lithuanian teachers' trade union | public |
| | lsadps | Lithuanian health workers' trade union | public |
| | lsmpsf | Lithuanian Education and Science Trade Union Federation | public |
| | lsso | Lithuanian Nursing Specialists Organization | public |
| | LTF | Lithuanian Transport Federation | public |
| | ltpf | Lithuanian law enforcement officials of the Federation | public |
| | lvpf | Lithuanian public service trade union federation | public |
| | valstybestarnautojai | Lithuanian civil servants, budget and public institutions darbuotojųprofesinė Union | public |
| Luxembourg | LCGB | Confederation of Christian Unions in Luxembourg | both |
| | OGB-L | Confederation of Independent Trade Unions | both |
| | ALEBA/UEP-NGL-SNEP | White-collar Union Federation | private |
| | FNCTTFEL | National Federation of Railroad Workers, Transport Workers, Civil Servants and Employees | private |
| | CGFP | GEneral Confederation of the Civil Service | public |
| | FGFC | Federation of the Municipal Administration | public |
| Malta | CMTU | Confederation of Malta Trade Unions | both |
| | GWU | General Workers' Union | both |
| | UHM | Malta Workers' Union | both |
| | MDU | Malta Dockers Union | private |
| | PSEU | Professional and Services Employees Union | private |
| | MAM | Medical Association of Malta | public |
| | MUMN | Malta Union of Midwives and Nurses | public |
| | MUT | Malta Union of Teachers | public |
| Netherlands | CNV | The National Federation of Christian Trade Unions in the Netherlands | both |
| | FNV | Dutch Trade Union Federation | both |
| | MHP | Federation of Managerial and Professional Staff Unions | both |
| | CNV Dienstenbond | Services Union | private |
| | CNV Vakmensen | Industrial union | private |
| | FNV Bondgenoten | Industriy, agriculture and services | private |
| | FNV Bouw | Construction, painters and woodworkers | private |
| | FNV Kiem | Artists, media, information and gaming | private |
| | AbvaKabo FNV | Civil servants, healthcare, semi-government and energy | public |
| | AOb | Education | public |
| | CNV Onderwijs | Teachers' Union | public |
| | CNV Public | Civil service and health care sector | public |
| Norway | LO | Norwegian Confederation of Trade Unions | both |
| | NUMGE | Norwegian Union of Municipal and General Employees | both |
| | Unio | Confederation of Unions for Professionals | both |
| | YS | The Confederation of Vocational Unions | both |

| Country | Abbreviation | Name, description | Type |
|---|---|---|---|
| | EI & IT | The Electrician and IT Workers' Union | private |
| | NJF | Norwegian Union of Railway Workers | private |
| | NTF | Norwegian Transport Workers' Union | private |
| | NTL | Norwegian Civil Service Union | public |
| Poland | fspolem | The Federation of Trade Unions of Cooperatives, Production, Trade and Services in Poland | both |
| | FZZ | Trade Unions Forum | both |
| | NSZZ | Independent and Self-governing Trade Union – Solidarity | both |
| | NSZZ | Independent Self-Governing Trade Union "Solidarnosc-80" | both |
| | NSZZS | Solidarity '80 \| both \| \| \| NSZZSK \| Solidarność'80 Trade Union Confederation \| both \| \| \| NZZK \| Drivers NZZ \| both \| \| \| OPZZ \| All-Poland Alliance of Trade Unions \| both \| \| \| Sierpień 80 \| Sierpień 80 \| both \| \| \| zzkontra \| TRADE UNION "VERSUS" \| both \| \| \| FZPTSLP \| Federation of Trade Union of Road Transport Workers of Telecommunications in Poland \| private \| \| \| FZZGWB \| Federation of Trade Unions Brown Coal Mining \| private \| \| \| FZZP \| Federation of Railway Workers' Unions | private |
| | PZZG | Alliance of Trade Unions Mining | private |
| | rkzzm | Confederation of Railway Trade Unions | private |
| | ZZGP | Trade Union of Miners in Poland | private |
| | ZZJG | Mining Trade Union Unity | private |
| | ZZMW | Machinists Union of exhaust Mine in Poland | private |
| | ZZPPMSPS | Workers Union of Meat and Food Industry in Poland seated HI "Heinz" Poland SA | private |
| | mzzpritv | The Federation of Trade Unions of Public Broadcasting in Poland | public |
| | fzzpozips | The Federation of Trade Unions of Health and Social | public |
| | MZZPOZ | Trade Union of Health Care Workers | public |
| | NSZZPP | Workers Trade Union of Police | public |
| | OZZFPSMG | National Union of Officers and Employees of Municipal Guard and the Municipal | public |
| | OZZPP | National Trade Union of Nurses and Midwives | public |
| | WZZSO | The Free Trade Union Solidarity-Education | public |
| | znp | Polish Teachers' Union | public |
| | ZZPEA | Trade Union of Workers Enforcement Administration | public |
| | zzpkm | Integration of Trade Unions of Public Transport in Poland | public |
| | ZZPKM | Workers Union of Public Transport in the Republic of Poland | public |
| | ZZPP | Police Workers Union | public |
| | ZZUP | Polish Health Trade Union | public |
| | ZZZSC | Association of Trade Unions of the Customs Service of the Republic of Poland | public |
| Portugal | UGT | General Workers' Union | both |
| | CGPT | General Confederation of Portuguese Workers – Intersindical | both |
| | SINAPE | Sindicato Nacional dos Profissionais da Educação (Education Union) | dont know |
| | SIFA | Independent Union of Railway and Allied | private |
| | SINAFE | Sindicato Nacional Ferroviários do Movimento e Afins (Railway Union) | private |
| | SINDEL | Sindicato Nacional da Indústria e da Energia (Union for Industry and Energy) | private |

| Country | Abbreviation | Name, description | Type |
|---|---|---|---|
| | SINDEQ | Sindicato Democrático da Energia, Química, Têxteis e Indústrias Diversas (Union for Energy, Chemistry, Textile and other Industries) | private |
| | QUOTES | Trade Union Association of Health Administrative Staff | public |
| | SINDEP | Sindicato Nacional e Democrático dos Professores (Teachers' Union) | Public |
| | SINTAP | Sindicato dos Trabalhadores da Administração Pública e de Entidades com Fins Públicos (Public Sector Union) | public |
| | SITAP | Independent Union of Workers of the Public Administration | public |
| Romania | CNSLR-Frăția | National Confederation of Free Trade Unions of Romania – Brotherhood | both |
| | CNSLR/BNS | National Confederation of Free Trade Unions | both |
| | CSDR | The Democratic Trade Union Confederation of Romania | both |
| | CSN Meridian | Meridian National Trade Union Confederation | both |
| | Liga SMVJ | League of Jiu Valley Miner Unions | both |
| | NTUC | Cartel Alfa | both |
| | Agrostar | National Federation of Unions of Agricultural, Food, Tobacco, And Related Services | private |
| | CNS "Cartel Alfa" | National Trade Union Confederation "Cartel Alfa" | private |
| | FSLI | Federation of Independent Trade Unions and Petrom | private |
| | ANOSR | National Alliance of Student Organizations from Romania | public |
| | FNSA | National Federation of Unions from Administration | public |
| | Publisind | Publisind | public |
| Slovakia | IOZ | Integrated Union | both |
| | KOZ SR | Confederation of Trade Unions of the Slovak Republic | both |
| | OZ KOVO | | both |
| | OZ prac. TV a športu | OZ prac. TV and sport in Slovakia | private |
| | OZZ | The trade union association of railwaymen | private |
| | SGI | SGI Posts and Telecommunications | private |
| | ZOE | Energy Trade Union | private |
| | OZ justície | OZ justice in Slovakia | public |
| | OZES | OZ prac. of Education and Science of Slovakia | public |
| | SGIPA | SGI public administration | public |
| | SOZZ | Slovak Trade Union of Health and Social Services | public |
| Slovenia | ZSSS | Association of Free Trade Unions | both |
| | KNSS | (smaller confederation) | dont know |
| | Konfederacija 90 | (smaller confederation) | dont know |
| | KSS Pergam | (smaller confederation) | dont know |
| Spain | CC OO | Trade Union Confederation of Workers' Commissions | both |
| | CIG | Galician Unions Confederacy | both |
| | LAB | Basque Union | both |
| | SAT | Andalusian Workers Union | both |
| | UGT | General Workers' Confederation | both |
| | ELA | Basque Workers' Solidarity | both |
| | FCM | Federation of Transport, Communication and Sea | private |

| Country | Abbreviation | Name, description | Type |
|---|---|---|---|
| | FCS | Federation of Construction and Services (FCS) | private |
| | MCA | Metal and Construction | private |
| | FETE | Federation of Education Workers | public |
| | FI | Federation of Industry (FI). | public |
| | FSP | Federation of Public Services | public |
| | FSS | Federation of Health (FSS). | public |
| Sweden | LO | Swedish Trade Union Confederation | both |
| | SACO | (organization for union covering graduate-level employees) | both |
| | TCO | The Swedish Confederation of Professional Employees | both |
| | Vårdförbundet | (Health Sector Union) | both |
| | IF Metall | | private |
| | Sveriges Ingenjörer | (Union for Engineers) | private |
| | Unionen | | private |
| | Lärarförbunde | (Teacher's Union) | public |
| | Vision | | public |
| Switzerland | SGB/USS | Swiss Federation of Trade Unions | both |
| | Travail-Suisse | Swiss Workers' Federation | both |
| | SEV | Union of Transport Employees | private |
| | Unia | | private |
| | VPOD | Swiss Union of Public Sector Employees | public |
| United Kingdom | GMB | GMB | both |
| | NUT | National Union of Teachers | both |
| | TUC | Trade Union Congress | both |
| | Unite | Unite the Union | both |
| | USDAW | Union of Shop, Distributive and Allied Workers | private |
| | UNISON | UNISON | public |
| | RCN | Royal College of Nursing | public |