# OPTED

## A repository of political party and interest group texts

**Zachary Greene, Christoph Ivanusch, Pola Lehmann, & Thomas Schober**

**Disclaimer**

This project has received funding from the European Union's Horizon 2020 research & innovation programme under grant agreement No 951832. The document reflects only the authors' views. The European Union is not liable for any use that may be made of the information contained herein.

**Dissemination level**

Public

**Type**

Report

**OPTED**
Observatory for Political Texts in European Democracies:
A European research infrastructure

# A repository of political party and interest group texts

**Deliverable D4.2**

**Authors: Zachary Greene[1], Christoph Ivanusch[2], Pola Lehmann[2], & Thomas Schober[1]**

[1] University of Strathclyde, Glasgow

[2] WZB Berlin Social Science Center

**Due date:** May 2021

# 1 Identifying Political texts from online sources

Work Package 4 of OPTED has created a repository of links to data collections of political text produced by political parties and interest groups (IG), using the definitions and criteria outlined in the first deliverable (D4.1). The repository can be accessed here: https://perma.wzb.eu/opted_wp4_inventory. The inventory consists of two spreadsheets (datasets). One spreadsheet each for political parties and IGs. It contains information about the corpora we identified including temporal, country, and party coverage, accessibility, and usability. The variable text.category informs the user about the text type that can be accessed via the specific corpus, e.g. press releases, manifestos, websites. We regard the repository as a living document that will continue to grow and be updated over the full course of the design phase. The codebook to this repository can be found here: https://perma.wzb.eu/opted_wp4_codebook_inventory. In this report, we provide an overview of our approach to identifying data repositories and the available data we have identified in key areas to date. In the process of creating this repository, we have also identified areas where data has yet to be compiled or is not readily available from online sources.

We see a number of *major gaps* in the available data that substantially limit scholars' ability to understand the role of political parties and IGs in democratic processes: (1) There are few systematic data collections of party and IG internal rules and proceedings. Without information on the internal processes of these organizations, it is difficult to rigorously evaluate their ability to perform the representative functions that prominent political theories ascribe to them (e.g. Mair and Katz 2002; Aldrich 2011). (2) Much of the available data collections we identified comes from more official and institutional settings. Particularly for IGs, the data is limited to their behaviors in relation to official state targets such as lobbying reports. It is difficult to uncover power dynamics within these organizations or between them when few official records exist outside of more formal environments. (3) Many available datasets suffer from a recency bias; data from current and recent events are much easier to find than older records from parties and IGs. This bias is particularly problematic for contexts such as organizational websites and social media profiles that contain past promises, indications of political priority and other valuable information at odds with the organizations' current leadership. Ultimately, the loss of historical data limits the formation of an accurate and shared public memory and constrains individuals' ability to hold organizations accountable for their past behaviors and messages. OPTED believes organizations have a responsibility to make data available for research and is ready to work as a strategic partner for a more systematic approach to data repository and storage.

# 2 Approach

Political parties and IGs produce a substantial amount of text in the public and private sphere. Despite the abundance of sources, much of this textual data has yet to be collated into comparable datasets. The WP4 team includes diverse experts with knowledge of research on many elements of both party and IG politics that served as instrumental in this collection, but is also biased in favor of knowledge of Western European countries. Consequently, we regard the inventory as a living document. The repository will continue to grow as new datasets become publicly available and as colleagues highlight datasets that are difficult to identify without country specific knowledge. OPTED in the long term will provide for a flexible set up to incorporate such sources into the inventory.

As a starting point, the team leveraged their case-specific knowledge to locate and identify large, cross-national datasets of textual data. Building on the definitions and areas of research identified in D4.1, the repository includes links to these large-scale data projects such as the Comparative Agendas Project and the Manifesto Project (e.g. Volkens et al. 2020). Political foundations such as the Friedrich-Ebert-Stiftung and the Fondation Jean Juarés contain a further wealth of political text related to parties' communications. However, data from these sources will require greater effort to produce meta data from that can be linked to broader cross-national datasets, because the structure of many foundations' archives are not set up to easily connect individual documents or collections to large cross-national databases. We also identified a number of national level archives with great potential for expanding the breadth of the datasets, but we require greater country specific knowledge and language skills to meaningfully exploit these sources.

Where we found no obvious pre-existing cross-national projects, we performed a country-focused

search to identify the availability of systematic data collections and raw texts. To do so, we combined case specific knowledge from the WP4 team with keyword searches in each country. For documents related to internal party communications this approach proved instrumental. Individual researchers have undertaken data collection projects that are likely available in a mixture of public dataverse's (e.g. GESIS; the Harvard Dataverse; Greene and Haber 2016) or are not yet readily available for public consumption. To supplement these case specific collections, we also performed a search for the primary source data from parties' websites. Most party websites include a recent news page that contain a mixture of newsletters, speeches, press releases, policy papers and other documents that can, although not always, extend backwards 5-10 years.

IG text offered a greater challenge to identify than those for parties. Most cross-national collections center on the IG's official interactions with government such as lobbying behaviors. Other collections derive from their official press releases. After extensively searching, there is very little public text on their internal deliberations or decision-making. Given the broad range and types of IGs that exist cross-nationally, an inclusive definition of IGs makes it difficult to create a single list of all IGs in a country and to go through official websites in the same way that the team did for the list of parties. Consequently, the evidence for the IGs is limited more to the information derived from our expert knowledge. It requires a large-scale effort to systematically identify relevant IGs across European countries and compile data produced by these organisations.

## 3    Political Parties

### 3.1    Manifestos

Political scientists have long used party manifestos to study political parties' preferences and priorities (e.g. Budge and Farlie 1983). Consequently, an extensive database of manifestos are available from the Comparative Agendas Project (e.g. John et al. 2013; Alexandrova et al. 2014; Breunig and Schnatterer 2020; Baumgartner et al. 2019), the Manifesto Project (e.g. Volkens et al. 2020), and Polidoc (Benoit et al. 2009). A number of country-level projects also exist that collect and annotate manifestos using alternative frameworks and for different levels of parties (AUTNES; Regional Manifesto Project). For most parties in European countries, manifestos are available in machine readable format and in many cases paired with content annotations produced by either or both the Agendas Project and the Manifesto Project extending from the post-war era for most Western European countries and since the transition to democracy for Eastern European countries. These datasets also include substantial meta data linked to the elections held at the same time the manifestos were produced.

### 3.2    Party websites

Despite the growth of parties' online presence since the early 2000s (e.g. Scarrow 2018), there is no central database of political parties' websites in Europe. However, there is a wikidata page available that compiles the links to parties' websites. These pages could be scraped, but would likely not allow for much over-time comparison as pages are frequently updated with little announcement. Some parties' historical websites, including now-defunct parties, are likely available from the internet archive, although many existing organizations set up redirect tools to reduce the availability of their pages from the archive. The sites will need to be scraped to be part of a central database.

### 3.3    Statutes

Parties' statutes and rules have become of increasing interest to political scientists and practitioners alike. Consequently, organizations such as the Rosa Luxemburg foundation and the National Democratic Institute have created compilations of their rules at fixed points in time. Older attempts at noting parties' internal rules and procedures building on work by Katz and Mair's (1994) party organization dataset did not include the text of the rules. Likewise, the Party Rules Database (Poguntke et al 2016) has compiled links to many parties' rules, but has focused more on the official rules regulating parties by governments, rather than party generated texts. The formatted texts of party statutes are available from the Party Rules Database website for more recent years. The listing for the parties' statutes, therefore, primarily includes a set of links

to the project and foundation websites that can be used to crawl the current rules. Many of these statutes are likely also available directly from the parties' websites.

### 3.4    Party newsletters and Magazines

Parties' produce substantial information for their activists and supporters through their newsletters and magazines. Our search has not revealed an extensive collection of these documents. From personal experience (e.g. Greene 2012), we expect print copies of these to be held in parties' individual archive or their associated foundations. However, these are likely not fully public without extensive archival work. The extent to which they are publicly available online at the moment is often through parties' websites on their archival news stories. These vary by time and are intermixed with other texts including speeches and other news about the party.

### 3.5    Party conference speeches and motions

The Party Conference Research Group is currently collecting speeches and motions in a number of countries. The data associated with this project is largely still in the collection phase focused on approximately eight Western European countries. Data for France, Germany and Italy are available online in a collated database (Greene and Haber 2016; Ceron 2019) and will be available soon from Austria (Kaltenegger et al. 2019), Denmark and the Netherlands (Schumacher et al. 2019). The broader set of internal documents are either unavailable, particularly from more conservative parties, or have yet to be scraped from parties' websites. As the largest meeting of the party's organization in most countries, transcripts and important documents from these meetings are often available as they happen, but are then removed from the public's view shortly thereafter.

### 3.6    Public speeches

Politicians produce large numbers of speeches to diverse groups. Most of these are not available or compiled for public use. However, parties create extensive collections of selected speeches from key party members that they publish on party and candidate websites, often under the recent events page. We list the websites where these are available from a large selection of parties in our repository. The speeches will need to be scraped from the public sites. To be useful to a large community, the speeches will require meta data on the dates and context of the speech, as they sometimes are the texts of more formal settings and others are held with distinct constituencies.

### 3.7    Press Releases

Scholars have grown increasingly interested in the role of parties' and interest groups' press releases. Consequently, there are a number of projects that have collected them in individual countries. However, most of these datasets are not shared publicly, at least yet. There are some compilations of press releases from Austria, Germany and Switzerland in online dataverses (GESIS, Harvard), but the majority of available press releases are located at individual party foundations. A specifically rich source is the Friedrich-Ebert-Stiftung, which collects a substantial amount for a number of parties. In some countries, there may be a centralized press agency that collects and archives press releases, but at this point we have only identified press releases from the Austrian Press Agency. Despite increasing interest from scholars, our search suggests that the majority of this data would have to be scraped from these foundations to be added to the OPTED repository.

### 3.8    Social media

Social media has become a common tool for politicians to communicate directly with their constituents. Consequently, scholars have also focused on the text produced by parties and interest groups. However, there are relatively few public datasets collected that we identified. Searches through common projects and dataverses revealed social media data related to Austria, Germany, Spain, and for Members of the European Parliament based on selected time periods. There are likely to be individual data sets available for different

countries, time points and topical foci (e.g. Heidenreich et al., 2020), but these are unsystematic and often not for secondary use. The types of social media listed are also fairly limited to Twitter. More extensive collections would require directly scraping data from the original sources, limited by the possibilities offered by different social media platforms.

## 3.9    Coalition agreements

Relying heavily on the expertise of the WP4 team, there is a comprehensive dataset of published coalition agreements collected by Klüver and Bäck (2019). The data is largely collected from government websites and by contacting individual parties where necessary. The collection is comprehensive including all coalition agreements in Europe dating back to the 1940s in some cases. These will be incorporated into the repository when they are made publicly available.

# 4    Interest Groups

In contrast to our efforts for political parties, we were able to only identify a smaller selection of compiled textual documents from IGs. Likely due to the difference in regulations on their activities and requirements for transparency, there are fewer online repositories of their activities. We have been most successful in identifying official interactions with government agencies through consultations and press releases. Likewise, a comprehensive list of all IGs for a country is difficult to come by. We identified one such list from a register of European Union lobbying activities, but this is limited to the EU level. Through this list of IGs' lobbying behavior, we could identify a list of their broader organizations and websites to collect data from. However, this approach limits knowledge of the broader range of IGs in existence that might not directly engage in lobbying government agencies or politicians. This increases the requirement for case specific knowledge to undertake the data collection exercise. This may limit the depth of our search, particularly in countries outside of the team's case knowledge.

## 4.1    Interest group statutes, internal deliberations, newsletters and magazines

Our search of major dataverses, individual projects and IG sites revealed a limited amount of information about IGs' internal deliberations. As IGs are not regulated in the same way as parties in most countries, legal requirements to publish this information are difficult to come by. IG websites may serve as a fruitful venue for future collections.

## 4.2    Interest group public speeches

A search for IG speeches did not uncover any major datasets. A repository of IG datasets indicates that most major collections relate to the participation of IGs in the policy-making process or in lobbying contexts. This reveals some letters in Sweden (e.g. Naurin and Boräng 2012) and parliamentary agendas from IG interventions in a handful of countries.

### 4.2.1    Interest group press releases

Although public dataverses did not offer major datasets of IG press releases, foundations offer a greater potential. The Friedrich-Ebert-Stiftung provides a substantial time series of IG press releases. The selection of IGs associated with this foundation relates to social democratic parties and trades unions in the 25 countries they list data for. There are some additional selections of press releases from the EU and EU related IGs in Germany and France from the foundation. There is an extensive list of IGs and their press releases from Austria due to the existence of the Austrian Press Association. Overall, there is substantial potential to scrape these press releases from trade unions across the EU, but there are very few obvious alternative compilations of press releases for other IG types. A detailed search of foundations associated with parties from other historical backgrounds may reveal a greater diversity of press releases.

## 4.3    Interest group websites

Much like the challenges of creating a repository for political party websites, IG websites change

frequently and we face the additional challenge of lacking a comprehensive list of IGs. We uncovered no consistent data collections of their websites. Here, we benefit from documentation on the lobbying behaviors of IGs as a method for identifying a list of websites. In particular, the lobbyfacts.eu website includes a comprehensive list of IGs involved in lobbying for each country. These profiles not only provide substantial information over their financial activities, it also lists the websites from each country. This will likely be a time intensive endeavor to fully explore and scrape these sites. Each site is structured differently and likely includes varying types of information. However, this list creates a method for identifying the universe of IGs at least in relation to the EU. Klüver (2019) has created a similar list of IGs in Germany that could be used as a starting point for identifying their websites, although the final year of the data collection is 2014.

### 4.4    Interest group social media

Our search found few records of IG social media datasets. There is an ongoing project collecting some data to the Spanish Agendas Project, but it is not yet publicly available (Bonafont et al. 2015). This seems like a promising area of data collection, as many organizations maintain a social media presence that is likely more consistent than their website and other online listings.

### 4.5    Interest group consultations

Interest group consultations also revealed few rigorous collections of textual data. Both searches from the Harvard and GESIS Dataverses and a search of recent publications for replication data did not yield major results. Projects including the Comparative Agendas Project (Baumgartner et al. 2019) and INTEREURO (Klüver et al. 2015) revealed some evidence of consultations data in Austria, Germany, Spain and in relation to the European Union. The team relied very heavily on prior individual knowledge of specific cases to complete this collection and would therefore likely require country experts to lead the collection from parliamentary and government websites for a broader range of countries. These texts are difficult to identify through general key word searches, so greater case specific knowledge of websites and the specifics of government organizations would likely increase the collection

## 5   Summary

In this first attempt at creating a data repository of political texts related to political parties and interest groups, we uncovered substantial inequities in data availability. We summarize the contents of the current repository in Table 5.1. Political parties, likely due to transparency requirements and their need for long term reputations, place a substantial amount of information online at least since the early 2000s. IGs post substantially less information. In both cases, efforts to scrape available sources to put them in a common, easy to use framework would improve the usability of this text to scholars. This would be particularly beneficial for press releases and the news sites from most political party websites as this data is widely available, but not systematically collected. Further datasets in both cases may be revealed through replication datasets related to recent publications. A primary source for future development of these datasets will therefore be the creators of these projects themselves. This places substantial weight on the creation of a common infrastructure for collating and connecting the related datasets as proposed in the OPTED framework.

| Table 5.1 | Overview of text types and availability | | | |
|---|---|---|---|---|
| | Internal Communication | Availability | External Communication | Availability |
| **Parties** | *Party Congress Speeches & Motions* | 2 countries, 1950s-present | *Press Releases* | 30 countries + EU, 2010-present |
| | *Statutes* | Selected parties in most European countries | *Websites* | 2 countries, selected time periods |
| | *Newsletters & Magazines* | No identified collections | *Manifestos* | Extensive availability in multiple formats |
| | | | *Coalition Agreements* | There is a single known dataset, but it is not yet publicly available, extensive cross-country coverage |
| | | | *Public Speeches* | No known dataset available, but list of links in party archives |
| | | | *Social Media Posts* | Selected parties and time periods available for 3 countries and the European parties |
| **Interest Groups** | *Statutes, Newsletters, Magazines, and internal deliberations* | No datasets currently available | *Press Releases* | Extensive data for trades unions across EU countries, but limited availability of other IG types |
| | | | *Website* | 6 countries for current periods + EU related lobbying pages for IGs in all EU countries |
| | | | *Position Papers* | No known dataset available |
| | | | *Consultation Submissions* | Extensive availability in 3 countries + the EU for periods in the 1990s-2000s |
| | | | *Public Speeches* | Limited Availability in various countries related to parties |
| | | | *Social Media Posts* | Limited availability for IGs in Spain |

# References

Aldrich, JH. 2011. *Why Parties?: A Second Look*. Chicago University Press.

Alexandrova, P; M Carammia, S Princen, and A.Timmermans. 2014. "Measuring the European Council Agenda: Introducing a New Approach and Dataset." *European Union Politics* 15(1): 152–67.

Baumgartner, FR, C Breunig, and E Grossman. 2019. *Comparative Policy Agendas: Theory, Tools, Data*. Oxford University Press.

Benoit K, T Bräuninger, and M Debus. 2009. "Challenges for estimating policy preferences: Announcing an open access archive of political documents." *German Politics* 18(3): 440-453.

Bonafont, LC, FR Baumgartner, and A Palau. 2015. *Agenda Dynamics in Spain*. Palgrave Macmillan UK.

Breunig, C, and T Schnatterer. 2020. "Die Politische Agenda Deutschlands." *Springer*.

Budge, I, and D Farlie. 1983. *Party Competition: Selective Emphasis or Direct Confrontation?: An Alternative View with Data*. SAGE Publications Inc.

Ceron, A. 2019. *Leaders, Factions and the Game of Intra-Party Politics*. Routledge.

Greene, Z, and M Haber. 2016. "Leadership Competition and Disagreement at Party National Congresses." *British Journal of Political Science* 46(3): 611–32.

Greene, Z. 2012. *Motivating parliament: the policy consequences of party strategy*. PhD Dissertation Completed at the University of Iowa, Iowa City, USA.

Heidenreich, T, JM Eberl, F Lind, and H Boomgaarden. 2020. Political migration discourses on social media: a comparative perspective on visibility and sentiment across political Facebook accounts in Europe. *Journal of Ethnic and Migration Studies*, *46*(7), 1261-1280.

John, P, A Bertelli, W Jennings, and S Bevan. 2013. "Public Opinion and the Policy Agenda." In *Policy Agendas in British Politics*, Palgrave Macmillan UK, 130–52.

Kaltenegger, M, K Heugl, and WC Müller. 2019. "Appeasement and Rewards: Explaining Patterns of Party Responsiveness towards Activist Preferences." *Party Politics*: 135406881985420.

Katz, RS, and P Mair. 1994. *How Parties Organize: Change and Adaptation in Party Organizations in Western Democracies*. Sage Publications.

Klüver, H. 2019. "The Survival of Interest Groups: Evidence from Germany." *West European Politics* 43(7): 1436–54.

Klüver, H, and H Bäck. 2019. "Coalition Agreements, Issue Attention, and Cabinet Governance." *Comparative Political Studies* 52(13–14): 1995–2031.

Klüver, H., C Mahoney and M Opper 2015. Framing in context: how interest groups employ framing to lobby the European Commission. *Journal of European Public Policy* 22(4): 481-498.

Mair, P, and R Katz. 2002. *Political Parties: Old Concepts and New Challenges*. Oxford University Press.

Naurin, D, and F Boräng. 2012. Who Are the Lobbyists? A population study of interest groups in Sweden. *Statsvetenskaplig tidskrift* 114(1): 95–102.

Poguntke, T, SE Scarrow, PD Webb et al. 2016. "Party Rules, Party Resources and the Politics of Parliamentary Democracies: How Parties Organize in the 21st Century." *Party Politics* 22(6): 661–78.

Scarrow, SE. 2018. "Multi-Speed Parties and Representation: The Evolution of Party Affiliation in Germany." *German Politics* 28(2): 162–82.

Schumacher, G, D Hansen, MACG van der Velden, and S Kunst. 2019. "A New Dataset of Dutch and Danish Party Congress Speeches." *Research & Politics* 6(2): 1–7.

Volkens, A, T Burst, W Krause, P Lehmann, T Matthieß, N Merz, S Regel, B Weßels, L Zehnter. 2020. *The Manifesto Data Collection. Manifesto Project (MRG/CMP/MARPOR). Version 2020b. Berlin: Wissenschaftszentrum Berlin Für Sozialforschung (WZB). https://doi.org/10.25522/manifesto.mpds.2020b*.