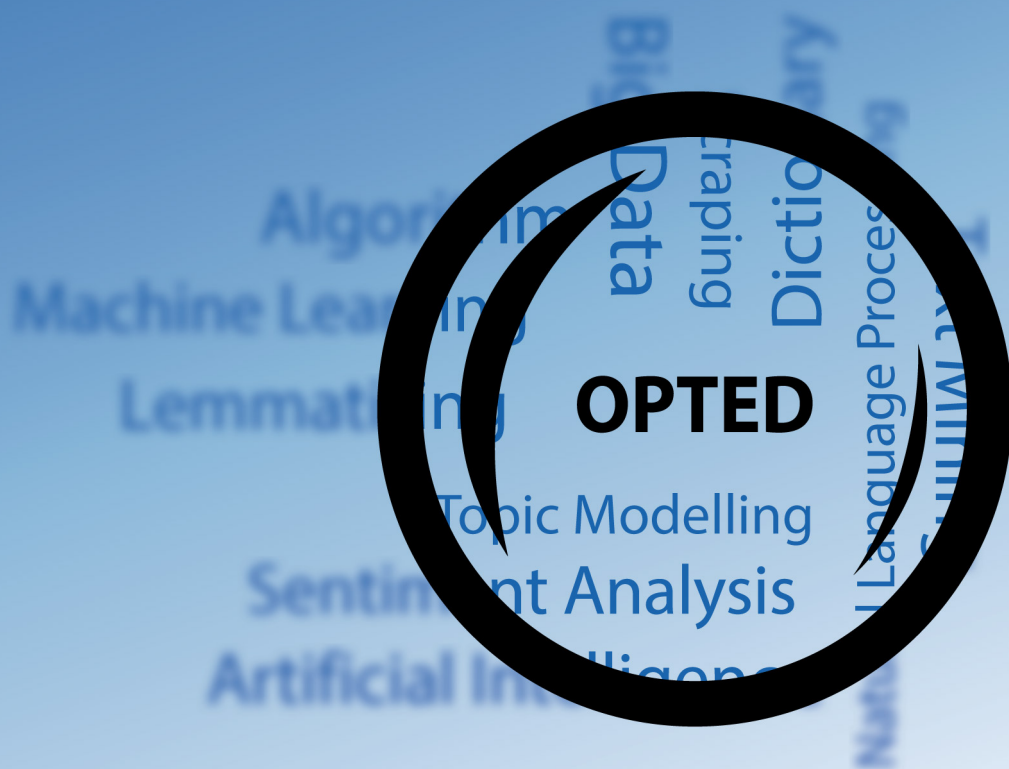


OPTED

Report on meeting bureaucrats/data scientists

Miklós Sebők, Sven-Oliver Proksch, & Christian Rauh



Disclaimer

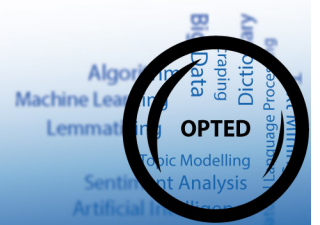
This project has received funding from the European Union's Horizon 2020 research & innovation programme under grant agreement No 951832. The document reflects only the authors' views. The European Union is not liable for any use that may be made of the information contained herein.

Dissemination level

Public

Type

Report



OPTED

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

Report on meeting bureaucrats/data scientists

Deliverable D5.4

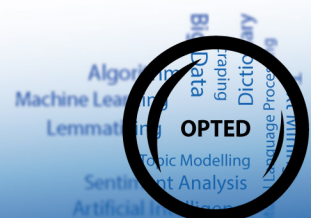
Authors: Miklós Sebők¹, Sven-Oliver Proksch², & Christian Rauh³

¹ Centre for Social Sciences, Budapest

² University of Cologne

³ WZB Berlin

Due date: June 2022



1 Narrative summary

Deliverable 5.4 tasks WP5 of the OPTED consortium to bring different stakeholder groups together that are interested in parliamentary data, from political scientists over data journalists and up to parliamentary authorities. On June 15th 2022, WP5 invited a diverse array of speakers to the European Parliament in Brussels to discuss the challenges and potentials of legislative text data. The following gives a summary of the event and more information on programme and speaker backgrounds.

1.1 Motivation

Parliamentary data provides a rich sense of information for political scientists. It can inspire research on inter- and intraparty dynamics, it allows to study long-term dynamics over multiple legislative sessions or the evolution of issues and ideas over long timeframes and comparative research on more general trends across the continent. These data sources might also enhance the transparency and understanding of political decision-making and legislative bodies for journalists and civil society organisations.

Yet research, both by academics and civil society, is still constrained by challenges of data access, data formats and a fragmented offering of services, where research questions might not be determined by substantive interest but the question of what can actually be studied. WP5 is working towards closing this chasm for the European research community by collecting resources, giving an overview of where data might be drawn from (D5.1), developing a new one-stop resource that combines multiple countries into one accessible dataset (D5.2) and furthering its dissemination with training materials (D5.5) and accessible interfaces (D5.6).

1.2 Planning

To ensure that these efforts reach a wider audience beyond the OPTED consortium of political scientists, but also other stakeholders such as NGOs or data journalists to WP5 organised a conference on “Parliamentary Data for Better Democracy” in the rooms of the European Parliament in Brussels on 15th June 2022. A website (<http://www.data4parliaments.org>) was set up to provide information on the conference and enable the registration for interested attendees, and announcements about the conference were distributed via Twitter, which reached 1178 impressions, according to Twitter Analytics.

In accordance with COVID protocols, participants were also offered the option of remote attendance, if they wished to abstain from travelling or had fallen ill. Using the facilities provided by the European Parliament, this allowed a smooth running of procedures and integration of onsite- and hybrid participation.

1.3 Speaker Contributions

WP5 selected a variety of guest speakers, according to different themes. Before the opening keynotes, Hajo Boomgaarden, the Project Coordinator of OPTED, introduced the consortium as a whole, welcomed participants and shared his enthusiasm for the event that day, highlighting how the work of WP5 ties into the wider goals and ambitions of OPTED.

The keynotes were begun by Christian Rauh, Jan Schwalbach and Sven-Oliver Proksch, all part of WP5 of OPTED, detailing their work on an inventory of legislative text data (D5.1) and their continued progress on a new dataset of speech and law data for parliaments, which will be released and detailed later in the project. Andrea Abellán talked about the DataJournalism.com project, detailing their resources for data scientists and research on the challenges they face when working with political data. Christian Heyer and Monika Jantsch from the German Bundestag parliamentary authorities detail the services the Bundestag provides for users and researchers interested in the legislative and deliberative process, what the introduction of an API meant, and their vision of how their data archive will evolve.

In the next session, dedicated to the potential of parliamentary text data to improve democracy through transparency, Attila Bátorfy began by detailing the journalism and visualisation projects he has been involved in at Átlátszó and Átló, building up a team dedicated to data journalism. Ieva Dunčikaitė from Transparency International Lithuania introduced the Manoseimas project, which gives citizens access to information about members of parliaments and their activities, and their lobbying work on furthering access to open data. Aiming to fight corruption through collecting and making legislative data openly available, Dominik Brenner highlighted the research of the CEU in the Global Corruption Observatory.

The session after lunch was dedicated to different data users. David Cabo from the Fundación Civio details how journalists and data scientists collaborate to collect different data sources and report on politics and media. Michal Ovádek, who maintains an R-package to access European legislative data, talked about the value of APIs for researchers and journalists and how data users might exploit more of their potential. Florian Richter and Jakob Kraus described how they make plenary debates of the German Bundestag available to researchers and the public via the Open Discourse platform.

The last topical session featured on the perspective of those authorities that hold and provide parliamentary data. Andrew Aiton from the Scottish Parliament shared how they use (text) data to support the work of parliament in public engagement, when working with public consultations for example.

Addie Erwin described the resources that the Inter-Parliamentary Union offers to researchers and data users, especially their PARLINE database on national parliaments worldwide. For the Austrian parliament, Christoph Konrath talked the perspective of parliamentary data providers on data availability and how they envision their future usage.

1.4 Outcomes

The programme featured stimulating discussions and fruitful debates. Some of the resources that were talked about were unknown or under-utilised, and gathered renewed attention from participants. Other exchanges allowed data users to speak directly with data owners, and share their common goals, but also diverging restrictions and perspectives that guide their work. Very valuable was the input from across the European continent, with participants sharing how their national contexts shape data availability and data use, and learning of some inspiring examples. The same was true for parliamentary participants, with many different institutions represented, both on the national but also the supranational level, that shared the different perspectives and systemic organisations behind the data access procedures in their institutions. Different backgrounds, from academia to journalism and NGOs, lead to participants exchanging ideas and working practises and what might one learn from another discipline.

Through the conference, the OPTED Twitter account was active in promoting the event and featuring the contributions by speakers. A collection of tweets reached 3712 impressions, combined with the engagement with partner accounts from the institutions of guest presenters. The OPTED website has also highlighted the conference in a news item, furthering the reach.

Further impressions can be found in the Social Media presence by the Chief Organiser, Miklós Sebök, here:

https://twitter.com/Miklos__Sebok/status/1536980342820716545,

https://twitter.com/Miklos__Sebok/status/1537030149899681792,

https://twitter.com/Miklos__Sebok/status/1537090044405501954

2 Official Programme

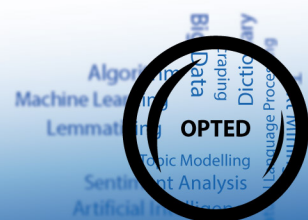
2.1 Welcome messages, Opening Plenary

Chair: Hajo Boomgaarden

Welcome messages: Miklós Sebök, Sponsors

This section featured keynote speeches and a discussion on “The Possibilities of Utilizing Parliamentary Data”. The concept of this section was based on the three key stakeholder groups of data journalists, parliamentary archives and academia. It aimed to provide an in-depth look into some successful, high impact project in each sphere. Presentations were 20 minutes each, followed by a discussion.

- *Academic perspectives:* **Miklós Sebök, Sven-Oliver Proksch, Christian Rauh, Jan Schwalbach** – Co-creators of the [OPTED inventory of legislative data](#), the [Parlawspeech database](#) and its interactive website



- *Data journalism perspectives*: **Andrea Abellán** – Datajournalism.com project of the European Journalism Center
- *Parliamentary archives*: **Christian Heyer, Monika Jantsch** – Directorate-General „Information and Documentation“ of the Administration of the German Bundestag
- *Discussion*

2.2 Roundtable: Better Parliamentary Data, More Transparency and Democracy?

Chair: Christian Rauh

This section featured a roundtable on the intersection between parliamentary data and democracy. The discussion focused on how more transparency regarding legislative speeches, procedures and outputs (such as laws) can contribute to a renewal of public discourse and representative democracy. The roundtable started with very brief presentations of the participants’ projects and institutions (10 minute each) followed by a roundtable discussion.

- **Attila Bátorfy** – ELTE University, Hungary, investigative/visual journalist at Átlátszó/ATLO
- **Ieva Dunčikaitė** – Manoseimas project, Lithuania
- **Martina Schories** – Data journalist, Süddeutsche Zeitung

2.3 Showcase of data users

Chair: Jan Schwalbach

This section featured a showcase of seven best practices of how to use parliamentary data. The line-up included presentations by legislative transparency projects, data journalism endeavours and other non-academic projects. Demonstrations present the challenges of procuring data in the right format, securing co-operation from data providers, data storage issues and convincing cases of visualization. Presentations were 15 minutes each, followed by a discussion of 30 minutes.

- **Dominik Brenner** – Central European University, Global Corruption Observatory
- **David Cabo** – Fundación Civio, an independent, non-profit newsroom in Spain
- **Michal Ovádek** – University of Gothenburg, Maintainer of the eurlex R package on European law
- **Florian Richter and Jakob Kraus** – Open Discourse project covering the debates in the German Bundestag

2.4 Roundtable of Data Providers

Chair: Sven-Oliver Proksch

This section featured a roundtable of representatives of data providers. The discussion focused on how data is made available by legislative archives or academic projects. Themes covered included, inter alia, data structures and format, best practices for opening up API access to parliamentary data, open data licenses. After the introduction (5 minutes) the roundtable started with very brief presentations of the participants’ projects and institutions in 5 minute each. The ensuing discussion was 40 minutes long, followed by a Q/A session.

- **Andrew Aiton** – Data visualisation expert at the Scottish Parliament
- **Addie Erwin** – Data officer at the Inter-Parliamentary Union
- **Christoph Konrath** – Research and Support in Parliamentary Matters Unit, Nationalrat of Austria
- **Rolandas Pyragis** – EP unit for Reception and Referral of Official Documents

2.5 Closing Plenary

Chair: Miklós Sebök

The closing plenary focused on gleaning takeaways from the conference. It served as an input and starting point for a future report legislative data transparency in Europe. This would consist of a summary of what participants see as approaches that work, steps to spread and co-operations to pursue. The structure of the section was as follows:

- A summary of the best practice findings of the conference
- Revisiting ParLawSpeech and OPTED as a platform for sharing open parliamentary data
- Future steps of the network
- Discussion

3 Speaker Biographies

3.1 Andrea Abellán

Andrea Abellán is the Data Coordinator of DataJournalism.com, a project by the European Journalism Centre. She works with a community of over 15,000 global members across multiple platforms. Andrea writes about, researches and maps out the latest and most relevant trends seen throughout the data journalism landscape. She has taken part in the production of The State of Data Journalism Survey, 2021, a project that analyses the state of the field in recent years. Andrea holds a Master in New Media and Digital Communication from Utrecht University. At the moment she is pursuing a PhD in data journalism at Miguel Hernández University. She focuses on the use of ethics and inclusion in data journalism, and examines the latest evolution of the field brought on by events such as the Covid-19 pandemic, the climate crisis and ongoing global conflicts. She has reported for different media outlets in Spain and Italy.

3.2 Andrew Aiton

Andrew is the Data Visualisation Manager at the Scottish Parliament Information Centre (SPICE). SPICE is Parliament's impartial research and information service. The data visualisation team is responsible for producing visualisations for SPICE and the parliament's Committees. Part of his role involves analysing responses to consultations received through the Parliament's digital engagement tools. This can involve analysis of the text of thousands of responses, often turned around to tight timescales. He is also leading on the Parliament's scrutiny-focused data science project, which is looking at how we can better collect, store, manage and use parliamentary data.

3.3 Attila Bátorfy

Attila Bátorfy is a master teacher of data journalism and visualization, and Phd-student at the Media Department of Eötvös Loránd University, Budapest. He worked previously as a journalist at several Hungarian news outlets, most notably he was the head of data and visual journalism at investigative journalism center Átlátszó. Currently he is running the first Hungarian visual journalism project, Átló. He also serves as editor of Médiakutató (Journal of Media Research). He created many government-transparency and media literacy projects in collaboration with Central European University, Transparency International Hungary, Mérték Media Monitor, Center for Independent Journalism, and Association of Hungarian Content Providers. For his journalistic work he was awarded with the Transparency-Soma Prize, the Quality Journalism Prize and the Eörsi János Memorial Award. His research interests include comparative media system theories, political influence in the media and theoretical and historical aspects of data visualization.

3.4 Hajo Boomgaarden

Hajo Boomgaarden is the Project Coordinator of OPTED. He is Professor of Methods in the Social Sciences with a focus on Text Analysis at the [Department of Communication](#), University of Vienna. He received his PhD in Communication Science from the [University of Amsterdam](#) in 2007. Much of his work revolves around the analysis of (news) media coverage and its effects on political cognitions, attitudes and behaviour. He has co-authored more than 100 articles in international journals and numerous book chapters. In his current position Hajo Boomgaarden deals with further advancements of computational methods, in addition to his ongoing work on political communication.

3.5 Dominik Brenner

Dominik Brenner is a postdoctoral fellow at the Department of Public Policy at Central European University.

His research interests include legislative favoritism, regulatory predictability, political oversight mechanisms, and the international political economy of finance. He is currently involved the OSUN-funded Global Corruption Observatory which collects legislative data, develops regulatory predictability measures, creates corruption indicators and probes the effects of legislative favoritism on company performance in countries around the world.

3.6 David Cabo

David Cabo is the co-director and CTO of Fundación Civio, an independent, non-profit newsroom based in Spain that develops impactful journalism projects in the public interest, focusing on public policies and their impact on citizens. Civio's work has been recognised internationally (e.g. Data Journalism Awards for Best Investigation and Best Individual Portfolio, García Marquez Innovation Award), and part of networks such as GIJN or the European Data Journalism Network. MSc in Software Engineering and BA in Psychology, David worked previously as a consultant, developer, and software architect for companies such as British Telecom, HM Revenues and Customs, Accenture, Ericsson and BBVA Global Markets. He is an expert in open data, public data analysis and budget data.

3.7 Ieva Dunčikaitė

Ieva Dunčikaitė is leading various Transparency International Lithuania initiatives on open data for transparency and anti-corruption including political integrity initiative www.manoseimas.lt that recently was selected as one of the global champions in World Summit Awards 2022. She also works with initiatives related to democratic participation, access to information and good governance in the public and private sectors. Ieva holds a Master's of European Politics and a double Bachelor's degree in International Relations and French from University of Exeter, UK. She also studied in Sciences Po Paris, France and Catholic University of Louvain, Belgium.

3.8 Addie Erwin

Addie Erwin is the Data Officer at Inter-Parliamentary Union responsible for ensuring the PARLINE database (data.ipu.org) is the reference point for authoritative and up-to-date data on national parliaments. She contributes to coordinating data collection, quality control and management of data in Parline, the IPU's Parline Correspondents' Network, monitoring and reporting on SDG indicators 5.5.1a and 16.7.1(a), and ongoing technical developments and promotion of the database. She joined IPU in 2017 after working as a Policy Analyst at the Organisation for Economic Cooperation and Development (OECD) where her work focused on public governance and women's representation in public life.

3.9 Christian Heyer

Ministerialdirigent Christian Heyer is a Deputy Director General of the Directorate-General „Information and Documentation“ of the Administration of the German Bundestag. He is responsible for the Parliamentary

Library, the Parliamentary Archive, the Parliamentary Documentation and the Press Archive. He is also in charge of the Lobbying Register for the representation of special interests vis-à-vis the German Bundestag and the Federal Government, which provides all information also in machine-readable form. Furthermore Mr. Heyer is tasked with coordinating the general Open Data Strategy for the German Bundestag.

3.10 Monika Jantsch

Monika Jantsch is the Head of the Parliamentary Documentation Division within the Administration of the German Bundestag. She studied law and passed the second state examination, thus qualifying as a lawyer and for judicial office. However, her interest in constitutional law matters sent her on a career path towards the Administration of the German Bundestag. Prior to her role in the Parliamentary Documentation Division, she also worked in sections dealing with petitions, matters of the Rules of Procedure and the legal status of Members, and budgetary law. She was also Head of the Secretariats of the Commission for Children's Concerns and the Committee on Family Affairs, Senior Citizens, Women and Youth. Monika Jantsch is a co-author of the commentary on the Members of the Bundestag Act, published in 2002 by Walter de Gruyter.

3.11 Christoph Konrath

Christoph Konrath is a constitutional expert and political scientist in the Austrian Parliamentary Administration. He's the head of the unit 'Research and Support in Parliamentary Matters' in the Legal, Legislative and Research Services. Further on, he's the coordinator of the Area of Interest 'Parliamentary Practice and Procedure' of the European Centre of Parliamentary Research and Documentation (ECPRD) and one of the editors of the International Journal of Parliamentary Studies.

3.12 Jakob Kraus

Jakob Kraus is a Computer Science B.Sc. Student at the TU Berlin, working as a Full-Stack Engineer at Limebit and as the Lead Developer for the OpenDiscourse Project, an initiative that offers citizens interactive full-text access and analysis of debates in the German Bundestag, a crucial contribution to the transparency of the German State. In his free time, he uses his programming skills to create real time digital art on his oscilloscope and is also active in the open source community.

3.13 Michal Ovádek

Michal Ovádek is a postdoctoral researcher at the Department of Political Science, University of Gothenburg and incoming Lecturer in European Institutions and Politics at the University College London. His research addresses various facets of the interplay between law and politics in the European Union. His work has appeared in, among others, the Journal of European Public Policy, European Union Politics and the Journal of Common Market Studies. As part of his research, he has developed and maintains an R package for facilitating access to EU laws and policies using database APIs and he is interested in the promotion of open data policies. Michal Ovádek obtained his PhD at KU Leuven in 2020 with a dissertation on the institutional contestation of EU competences. Between 2019 and 2021, he worked as a political adviser in the European Parliament.

3.14 Rolandas Pyragis

Rolandas Pyragis works as administrator in the Unit for Reception and Referral of Official Documents in the Directorate for the Plenary, Directorate-General for the Presidency (DG PRES), at the European Parliament. He is responsible, among other things, for managing the Legislative Observatory, the European Parliament's tool for monitoring the EU decision-making process, with a particular focus on Parliament's activities. Prior to joining DG PRES, he worked as a press officer and web editor at the Directorate-General for Communication. He studied international relations and political science at the University of Vilnius. Although more comfortable with the subjects of social sciences, he also likes to keep on learning about tech things, such as data management.

3.15 Florian Richter

Florian Richter is Founder, Managing Director, and Data Scientist at Limebit GmbH – a Berlin based data science consultancy. He has an academic background in political science and sociology. Together with his team, he started the Open Discourse project – the first fully comprehensive and machine-readable corpus of the parliamentary minutes of the German Bundestag. In December 2020, the Open Discourse project was published as an open source database along with a website that provides interactive tools to research and explore all plenary speeches held in the Bundestag since 1949.

3.16 Martina Schories

Martina Schories is an expert in programmed research and data visualization. She is an award-winning data journalist (Nannen Preis and Reporterpreis, among others) and was covering political issues at Süddeutsche Zeitung. Currently she is working at University of Siegen bringing data (methodology) and design skills to the Digital Media and Methods team and the CRC “Transformations of the Popular”. She considers herself as a translator between (a small fraction of) the real world and code and vice versa. Her main means of communication are data visualizations.

