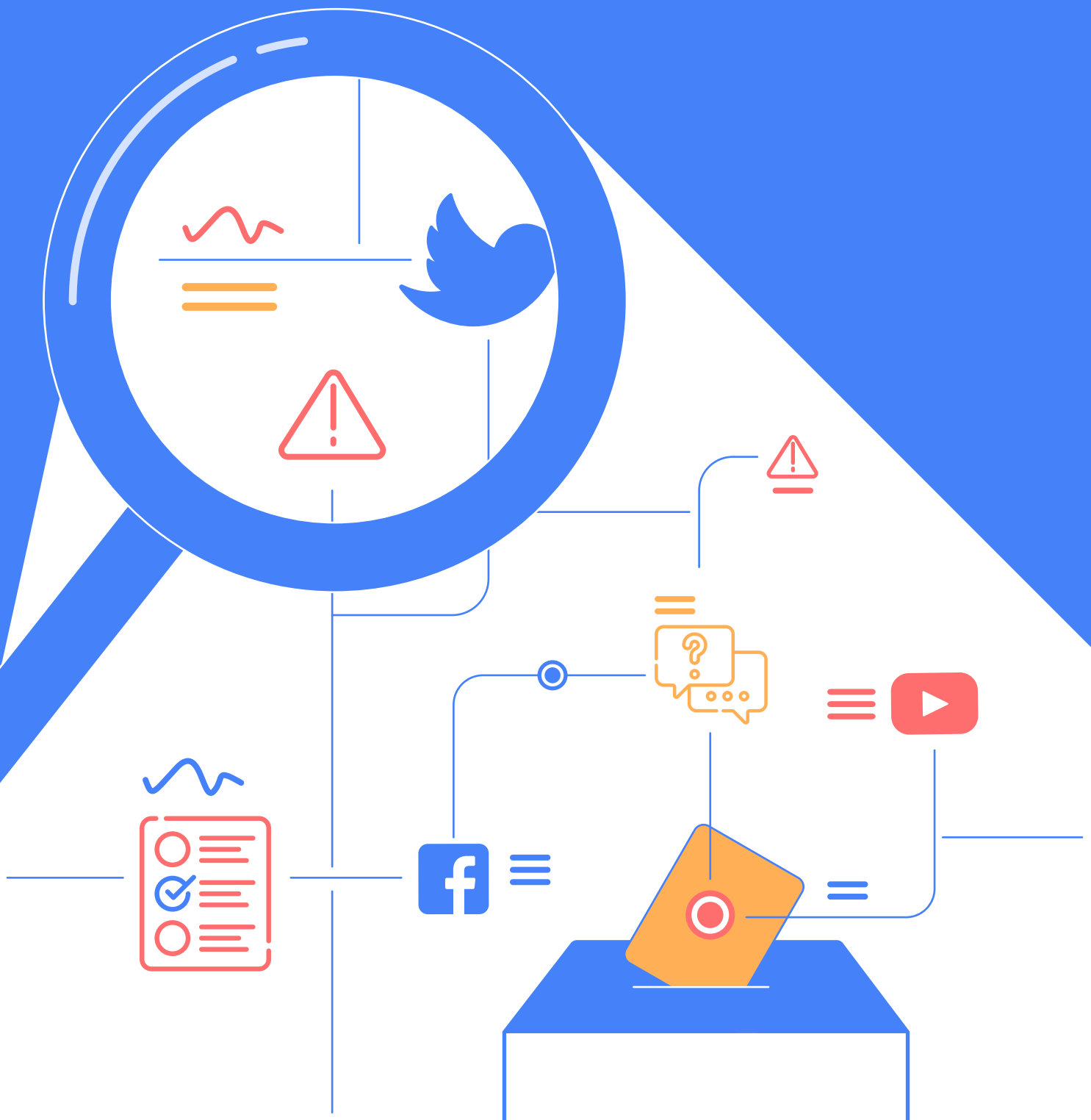


Lessons Learned:

Social Media Monitoring during Elections

Case Studies from five EU Elections 2019-2020



Lessons Learned: Social Media Monitoring during Elections

Case Studies from five EU Elections 2019-2020

Funded by:



With contributions by:



This publication is a result of a project carried out by Democracy Reporting International with contributions from MEMO 98 and the financial support of NEF-Civitates. The contents of this publication do not reflect the position of NEF-Civitates.

Under this project, five social media monitoring projects were carried out by teams from Election-Watch.EU (Austria), Lisbon University Institute (ISCTE - IUL) – Media Lab (Portugal), Political Accountability Foundation (Poland), Global Focus (Romania) and Gong (Croatia) during elections in the EU between 2019 and 2020.

Representatives from the five participating project teams convened on 20 March to discuss their social media monitoring activities. The presented approaches, lessons learned and recommendations have been summarised in this publication.

June 2020

Author: Madeline Brady



This publication is available under a Creative Commons Attribution Non-Commercial 4.0 International license.

Table of Contents

1.INTRODUCTION	4
2.CASE STUDIES: SMM APPROACHES FROM FIVE EU COUNTRIES	5
2.1 Austrian Legislative Election 29 September, 2019	6
2.2 Portuguese Legislative Election 22 November, 2019	7
2.3 Polish Parliamentary Election 13 October, 2019	8
2.4 Romanian Presidential Election 10 November, 2019	10
2.5 Croatian Presidential Election 22 December, 2019 and 5 January, 2020	11
2.6 Case Study Comparisons	12
3.LESSONS LEARNED TO HELP FUTURE SMM TEAMS	14
Preparation and Sample Selection	14
Data access, Collection and Analysis	15
Publication and Outreach	16
4.RECOMMENDATIONS TO HELP FUTURE SMM EFFORTS	17
Government	17
Social Media Companies	17
CSOs and Universities	18

Executive Summary

- There is no golden social media monitoring approach – projects must be context, organisation and resource specific
- Social media monitoring teams require both local political science expertise and data analysis skills
- Data access barriers are the largest hurdle, and data availability determines the reporting output
- Outreach activities help organisations maximise their impact by reaching policymakers and citizens
- After implementing social media monitoring once, organisations had the long-term infrastructure for continued work

1. Introduction

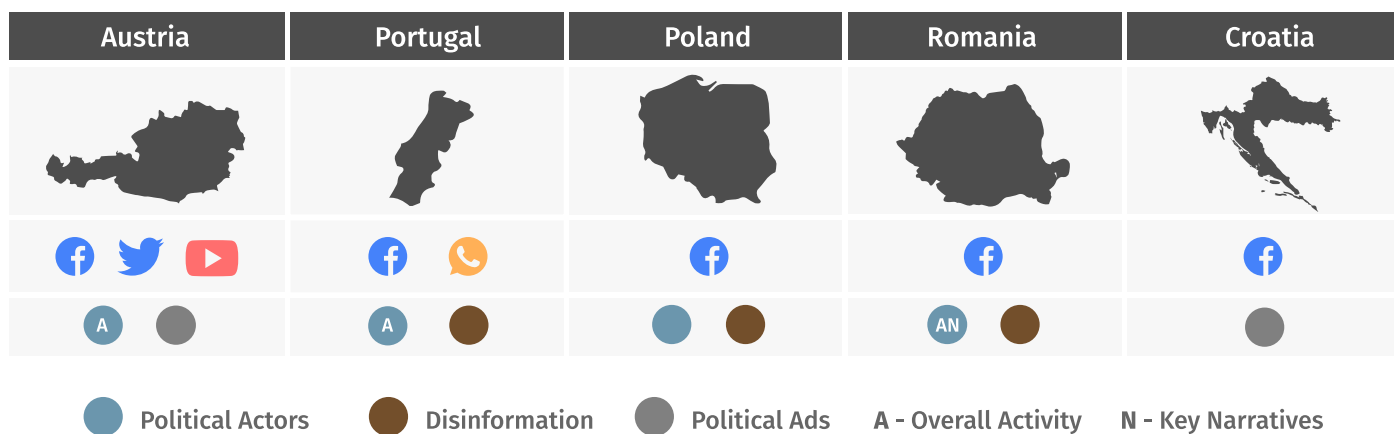
Social media platforms are important places for citizens to gather information, share their opinions and debate with others. As seen in recent elections around the world, malicious actors are spreading false, misleading or inflammatory content. Their goal is to confuse, polarise and ultimately undermine citizen’s belief in democratic systems. As emphasised by the OSCE Office for Democratic Institutions and Human Rights, democratic elections are a “key pillar of long-term security and stability”.¹ **Given the imminent threats posed to democratic discourse, monitoring social media platforms during elections has proven increasingly necessary.**

Social media monitoring (SMM) is the objective analysis of democratic discourse and political actors on social media platforms. As described in Social Media Monitoring During Elections written by DRI, monitoring social media is far more complex than monitoring traditional media “with a myriad of actors and content” ranging from official democratic institutions (e.g. parties, politicians, media) and unofficial actors (e.g. individuals, political influencers, partisan groups).² Additionally, SMM requires access to social media data, which may be difficult to access, limiting in terms of metrics and require technical data analysis skills. As a result, there are a number of different possible approaches that social media monitors may use.

So, what is the best way to monitor social media? As seen while monitoring five European elections during 2019 and 2020, the answer to this question must be context, organisation and resource specific. This field is still experimental in many ways and unfortunately, there is not one golden approach.

To help future teams get started with their monitoring activities, **this publication provides an overview of SMM analyses conducted in Austria, Portugal, Poland, Romania and Croatia.** More specifically,

- SMM approaches and key findings from each country
- Lessons learned from SMM
- Recommendations for government, platforms and research institutions to improve SMM



¹OSCE ODIHR, Election Observation Handbook, Warsaw, 2010, p.13, <https://www.osce.org/files/f/documents/5/e/68439.pdf>
²Goldzweig et al, Social Media Monitoring During Elections. Cases And Best Practice To Inform Electoral Observation Missions, Open Society Foundations, 2019, p.27, <https://democracy-reporting.org/wp-content/uploads/2019/07/social-media-monitoring-during-elections-20190614.pdf>

2. Case Studies: SMM Approaches from Five EU Countries

Five cases from elections in Europe between 2019 and 2020 indicate that this field is still experimental. Applied research on the role of social media during elections is important to complement academic and more long-term research on phenomena like the effect of disinformation and hate speech on political behaviour, and for its potential to capture current developments and hold tech companies more accountable. Also, accessing data while events unfold helps to preserve data that might be deleted afterwards. Since this is a new field, **there is not one single type of organisation, team makeup or methodological approach required for social media monitoring.**

Table 1 shows that teams were mixed with different experiences and skill sets. However, most teams included a combination of both social scientists and data analysts. They included between four and seven researchers.

Many of these organisations came from an election observation background and were seeking to incorporate a social media aspect into their electoral observation work. Other teams came from an open-source intelligence perspective looking to identify disinformation actors. Given their traditional media monitoring background, few of these teams used programming but rather programmes like Excel to analyse their data.

Table 1: Comparison of Organisation/Team Structure

Country	Organisation	Type of Organisation	Number of researchers	Team qualifications	Trad. Election Monitoring Experience	First SMM experience	Programming used**
Austria	Election-Watch.EU and VDSG/data4good	Election Monitoring NGO	7	Election observers, Political scientists, Data scientists, Social media experts	Yes	Yes	Yes
Portugal	Lisbon University Institute (ISCTE - IUL) – Media Lab	University Media Lab	4	OSINT* specialist, Journalist, Media studies	No	No	Yes
Poland	Political Accountability Foundation	Election Monitoring NGO	6	Election observers, social scientists	Yes	Yes	No
Romania	Global Focus	International Studies Think Tank	5	Media specialist, Journalist, Political scientists	Yes	No	No
Croatia	Gong	Election monitoring CSO	5	Media specialists, journalism, political finance expert and data scientist	Yes	No	No






* Open Source Intelligence (OSINT) is the collection and use of public data for intelligence purposes

** Programming meaning the use of computer languages (e.g. Python, Java, R)

The next section will deep dive into the five different approaches used by the teams in order to help future SMM teams get started. Each case study includes the following:

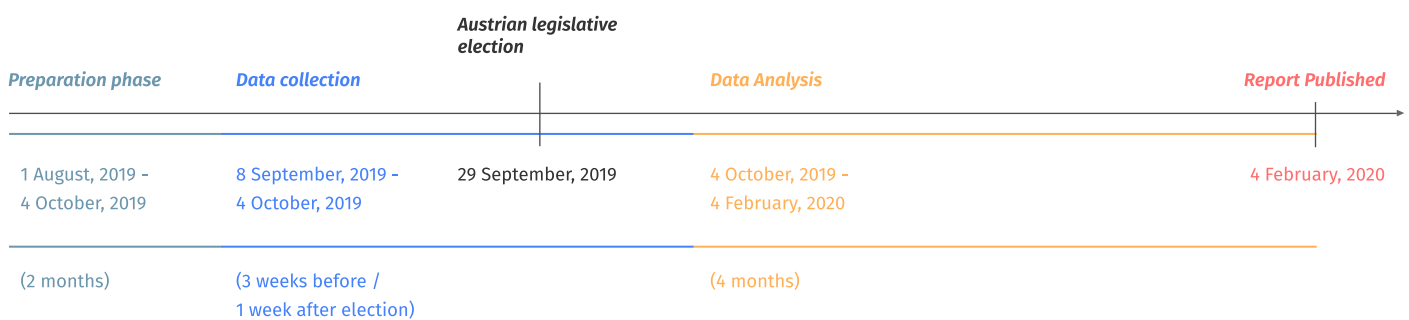
- Preparation and sample selection
- Data access, collection and analysis
- Publication and outreach

2.1 Austrian Legislative Election | 29 September 2019

    Political Actors  Political Advertisements **146 accounts | 25,000 posts | 34 days**

The Austrian snap legislative election (29 September 2019) was monitored by the election monitoring non-governmental organisation (NGO), Wahlbeobachtung.org in close cooperation with the Vienna Data Science Group– Data4Good and the Center for Applied Social Research of the Faculty for Public Policy Analysis of the University Fundação Getulio Vargas of Rio de Janeiro in Brazil (FGV DAPP). The team of seven included election observers, political scientists, data scientists and social media experts.

The team monitored **how official political actors (i.e. politicians, parties and media outlets) used Facebook, Twitter and YouTube during the political campaign.** In particular, they looked at how campaign activities (e.g. messaging, ads) changed over time and analysed how actors used different platforms. Ultimately, the team collected 25,000 posts and more than 1.1 million user comments from 146 different social media accounts.



Preparation Phase

The team spent two months identifying actors and keywords prior to the election. They conducted **pre-monitoring** from 28 July – 30 August to better understand the Austrian social media landscape.

For its **sample selection of “political actors”**, the team collected data from (a) candidates to the national assembly (b) political parties at the national and regional level (c) official partner organisations or support pages of political parties (e.g. youth organisations, unions, party TV channels) (d) active non-candidate users (e.g. former vice-chancellor) and (e) media outlets or journalists with more than 50,000 followers.

Data Access, Collection and Analysis

The team accessed data using the Facebook API³, Facebook Ad Library, Twitter API and YouTube API. Gaining access to the Facebook API took three months⁴, which proved to be quite challenging, unlike other platforms. The team also faced API stability issues with a number of API changes during monitoring.

For all platforms, the team analysed aggregate interactions with the political actors. For Facebook this included reactions, shares and comments on posts in addition to Facebook Ad expenditures. On Twitter this included favourites and retweets. On YouTube analysed view counts for the 25 top political videos.

Python and R were used to analyse the results. Given the large amount of data and team’s data science capabilities, the team attempted to use artificial intelligence (AI) techniques to analyse comment sentiment and identify key topics, but the methods did not produce qualitatively meaningful results.

Publication and Outreach

The team published an [online tool](#), one [final report](#) and conducted outreach activities to share their findings. The online tool allows users to interact with the data themselves and evaluate specific actors of interest. In addition to hosting an event with over 40 participants, the Austrian Public Broadcaster Radio broadcasted a report about the project on its morning news programme, Morgenjournal.

³ An application programming interface (API) enables the accessing of certain aspects of a platform’s data related to requests received and responses sent.

⁴ They found no clear process to gain research access to data in a timely manner (see section 3 for more information). The reasoning for receiving limited Facebook API access versus full CrowdTangle access was not clear. Facebook API access only included pre-selected pages without full access to all public pages unlike CrowdTangle.

2.2 Portuguese Legislative Election | 22 November 2019



Political Actors

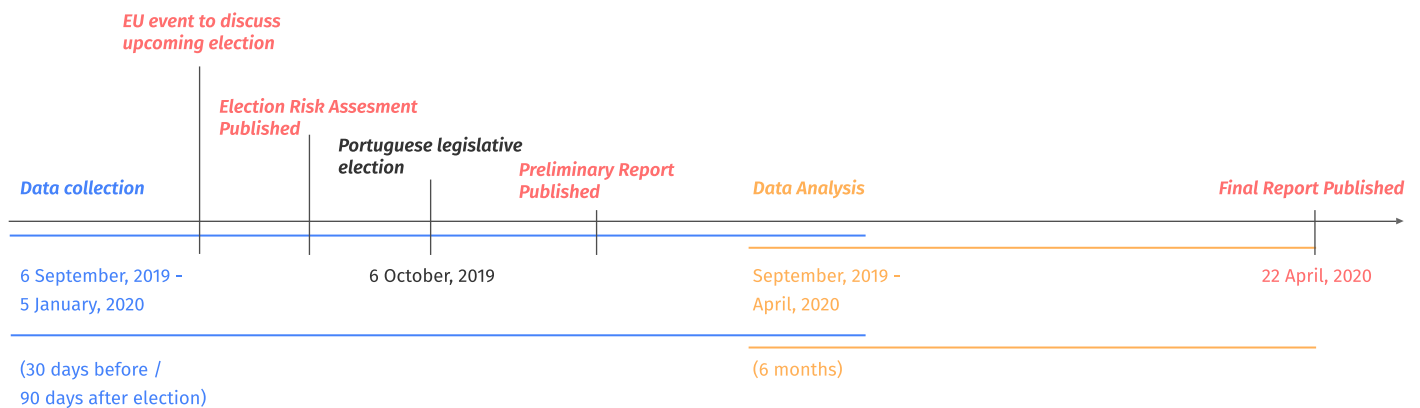


Disinformation

85 accounts | 138 000 posts | 120 days

The Portuguese legislative election (22 November 2019) was monitored by the Lisbon University Institute (ISCTE - IUL) Media Lab. The team of four included an expert in communication studies, a former journalist and an OSINT expert who was formerly a part of law enforcement.

The team monitored **disinformation-prone Facebook pages and groups** during the 30 days before the election and 90 days following the election (6 September 2019 – 5 January 2020). They also monitored **the evolution of newly elected political parties** to the new Portuguese parliament for growth and interaction with the traditional parties. The team collected 138,000 posts from 47 public Facebook pages and 38 Facebook groups.



Preparation Phase

In preparation, the team conducted a **risk assessment** analysing (a) social media consumption (b) trust in media (c) public attitudes and trust in institutions (d) electoral context and (e) political context.

For their **sample selection**, the team analysed accounts that frequently published posts about **corruption** via keyword search. Corruption was [previously identified](#) as a main concern for citizens regarding their political choices. After identifying non-mainstream political accounts discussing corruption, they analysed accounts above a threshold of followers, political posts and post frequency.



Data Access, Collection and Analysis

The team used **CrowdTangle** to access anonymised posts from the specific pages and groups. They analysed posts' text, links to external content and the publishing time via a self-developed tool. An off-the-shelf tool, **Foresight Brandwatch**, was used to identify key themes in posts and user comments.

Posts were manually labelled in line with the International Fact Checking Network's methodology according to content, context and visual clues and labelled (0) nothing to register (1) inaccurate (2) incorrect. Type of posts, discourse and theme were also labelled. They analysed the number of interactions these posts received, fan-base growth over time and the key disinformation narratives.



Publication and Outreach

Unique to this project, the **team partnered with a local news media organisation** and published findings in real time. Such an approach provided direct information to citizens who may have been impacted by online disinformation.

Besides news articles, **three reports** were published including a [risk assessment](#) prior to the election, a [preliminary report](#) two months after the election and a [final report](#) seven months after the election.

Additionally, with the support of DRI, the team panelled an **EU-hosted event** to discuss social media in the upcoming election. This event brought attention to the issue at the decision-maker level.

2.3 Polish Parliamentary Election | 13 October 2019



Political Actors

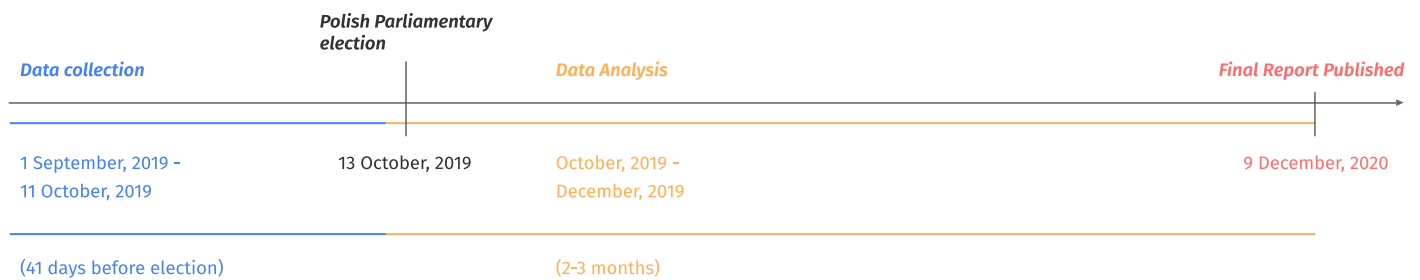


Disinformation

74 accounts | 8 848 posts | 41 days

The Political Accountability Foundation, an election observation NGO, monitored social media during the Polish parliamentary election (13 October) from 1 September to 11 October 2019. The team of six included traditional election observation experts in addition to social scientists conducting social media monitoring for the first time. The analysts underwent a two-day training to prepare for the project.

They analysed **how official political actors used Facebook during the election** to better understand the role of the country's most important social media platform. The team also studied the **key narratives** discussed by these actors and attempted to identify any disinformation narratives through a content analysis.



Preparation Phase

As the team was new to social media monitoring, a MEMO 98 expert conducted a **two-day training** to prepare analysts to for the project.

The team conducted a **risk assessment** by researching (a) social media consumption, trust in media (b) public attitudes and trust in institutions, and (c) electoral context.

For their **sample selection**, they identified a list of key political actors including parties, politicians, media and influencers based on qualitative research.

The team **pre-defined a list of preliminary narratives** based on qualitative research.



Data Access, Collection and Analysis

First, the team extracted post-level data using **NewsWhip**. This data included Facebook posts, post links, post messages, pictures, the time of posting and interactions (i.e. total likes, shares, comments, reactions). Second, they saved and analysed this data using **Microsoft Excel**.

They **manually coded** the content of each post based on the pre-defined list of narratives. If a post did not meet the criteria of the pre-defined list, coders assigned new codes. The final list included 65 narratives and included topics related to migration, nationalism and other social issues.

The team also analysed the **most active actor groups** and interactions to understand **user engagement**.



Publication and Outreach

The team published one [final report](#) following their monitoring activities, a little over two months after the election.

The results of the report were presented in an event with other election observation organisations and representatives of the National Election Commission.

2.4 Romanian Presidential Election | 10 November 2019



Political Actors

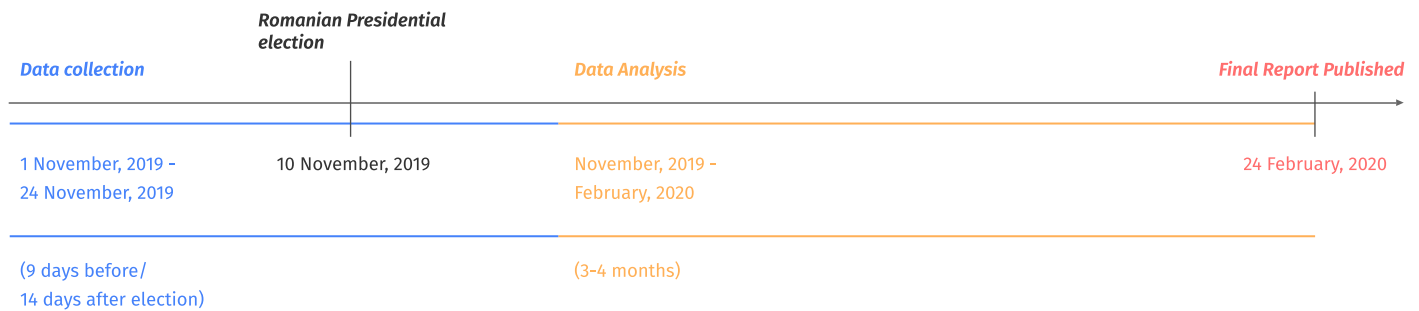


Disinformation

102 accounts | 14 507 posts | 24 days

Global Focus, an International Studies think tank, monitored Facebook during the Romanian Presidential Election (10 November 2019) from 1 to 24 November 2019. The team consisted of media specialists, journalists and political scientists.

The project studied the role and importance of Facebook during the election, as the country's most popular social media platform. They analysed **how official political actors used Facebook during the election** and analysed the **key narratives** pushed by these actors.



Preparation Phase

A MEMO 98 expert conducted an online training for Global Focus analysts to prepare them for the project. The team conducted a **risk assessment** by researching (a) social media consumption (b) trust in media (c) public attitudes and trust in institutions, and (d) electoral context. For their **sample selection**, they identified a list of the most important political actors including registered candidates, political parties, influencers/journalists and media outlets, based on qualitative research. They pre-defined a list of 100 **narratives** based on previous research the team had conducted earlier in the year during the European Parliamentary elections. This list included more general topics and issues (such as social issues, gender equality and education) as well as more local and specific narratives (such as the fight against corruption and anti-corruption protests).



Data Access, Collection and Analysis

The team extracted post-level Facebook data using **NewsWhip**, which included posts, post links, post messages, pictures, the time of posting, and interactions (i.e. total likes, shares, comments, reactions). They then saved and analysed this data with Microsoft Excel. They manually coded each post based on the pre-defined codebook of narratives and topics. New narratives that did not exist in the list were added, which resulted in a final total of **117 narratives**. Through this approach, the team was able to demonstrate that some actors were pushing false narratives to discredit their opponents. The team also analysed the **most active actor groups** and interactions to understand **user engagement**.



Publication and Outreach

The team published one in-depth [final report](#) a little over three months following the election. The results of the report were presented in an **event** with other election observation organisations, representatives of the Permanent Electoral Authority, journalists and politicians.

2.5 Croatian Presidential Election | 22 December 2019 and 5 January 2020

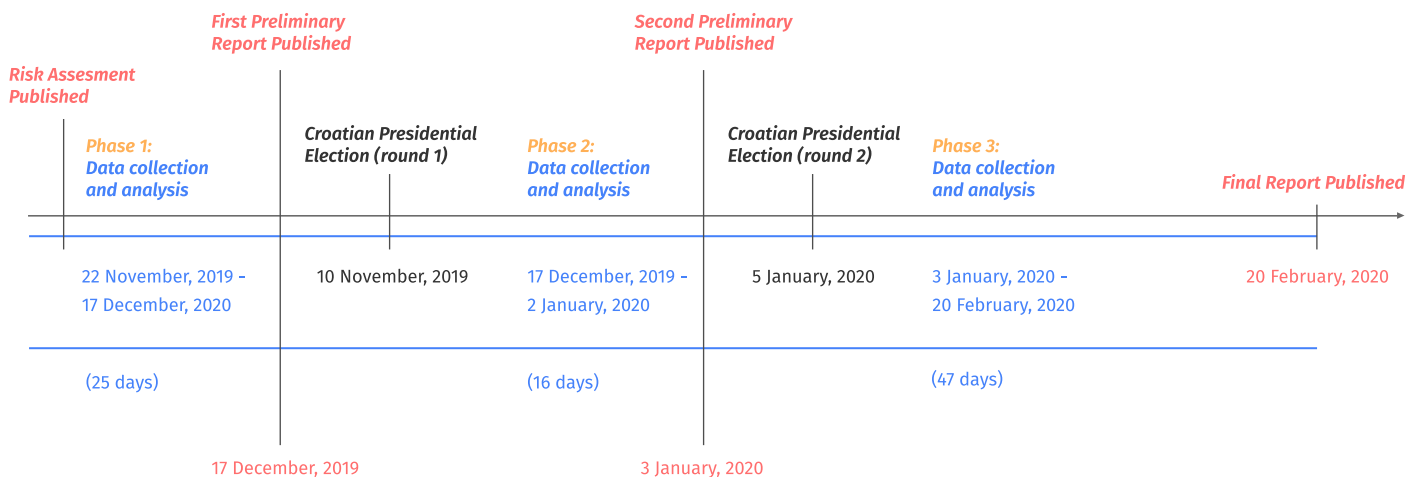


Political Advertisements

11 candidates | 3 financial reporting periods | 88 days

Gong, an election monitoring civil society organisation (CSO), monitored social media during both rounds of the Croatian Presidential Election (22 December 2019 and 5 January 2020). The first monitoring period took place from 22 November 2019 to 17 December 2019 and the second from 17 December 2019 to 20 February 2020. The team consisted of media specialists, journalists, a political finance expert and a data scientist.

A new campaign finance law requiring candidates to separately display their social media advertising costs in their financial reports was tested during this election. As a result, this monitoring project aimed to understand **whether politicians accurately reported political advertisement spends on Facebook** in their financial reports.



Preparation Phase

Prior to the election, the team conducted a risk assessment of the election by analysing (a) social media consumption (b) trust in media (c) trust in institutions (d) electoral context and administration (e) campaign finance on social media, and (f) political context.



Data Access, Collection and Analysis

The team retrieved data on each candidate's political advertising expenditures by using a combination of Facebook Ad Library, **Facebook Ad Library Report** and **Google's Transparency Report**. They also scraped data from candidates' official financial reporting published on the website of the State Electoral Commission (SEC). Data was analysed using **Microsoft Power BI** and **Tableau** was used for visualisations.

With both sets of information, the team analysed whether the reported social media advertising expenditure matched the candidate's official reporting. They experienced data limitations with Facebook's reporting tools, which made it hard to precisely compare Facebook expenditures to the financial reports.



Publication and Outreach

The team **directly reported errors in one candidate's financial reporting to the SEC**. This effort may have contributed to the candidate later correcting the error in their reporting.

Four real-time reports were published during this project. Prior to the election, a disinformation [risk assessment](#) was published. Two preliminary reports were published following the candidates' financial reporting during the [first](#) and [second](#) rounds of the election. Gong also published a [final report](#) with recommendations after candidates published their final financial reports.

Reports and findings were distributed through a number of **outreach activities** including **newsletters, press releases, Croatian media coverage**. The final report was presented at an **event** (Open Data Day conference) to a number of decision-makers at a multi-stakeholder conference.

2.6 Case Study Comparisons

A comparative overview of the five approaches can be found in Table 2.



Preparation Phase

In terms of setting up their activities, most groups conducted a risk assessment prior to their social media monitoring activities to gain a better overview of the election. Some teams were highly experienced with election observation (Croatia, Poland and Austria), while others had high technical capacities (particularly Portugal and Austria), which posed different challenges for different teams during the preparation phase. The Austrian team conducted pre-election monitoring to learn about the social media landscape and identify actors. Other teams were able to use external lists or previous institutional knowledge to identify social media accounts or narratives to study.

In selecting their samples, the teams attempted to understand social media during elections from different angles. Most of these studies looked at how the relevant electoral actors (e.g. politicians, parties, media and influencers) used social media (e.g. frequency of posts and likes) during the election and analysed the key narratives pushed by these actors. Other analyses (Croatia and Austria) looked at the question of political finance during elections by analysing political advertisements on Facebook. The Portuguese case zeroed in on known disinformation actors to analyse key narratives and techniques used by these accounts during the election.



Data Access, Collection and Analysis

Most teams looked at Facebook because it was the most popular platform in their country. They accessed data using CrowdTangle, Facebook API, NewsWhip or Facebook Ad Library/Report. The Portuguese team looked at WhatsApp via public group links and announced their presence in the group as researchers. The Austrian team looked at Twitter and YouTube using each respective platform's API.

Besides platform popularity, data availability and the technical capacities of the team helped define the scope of each project. Some studies looked at a wide range of platforms within a small time period, while others just focused on the most popular platform in their country for a longer timeframe. The Austrian team looked at the largest quantity of data and used Python to analyse and visualise the results. The other teams looked at a smaller amount of data and used Microsoft Excel to label and analyse the results.



Publication and Outreach

Some teams published one in-depth report several months following the election, while other teams published more frequently to provide insights to stakeholders during the campaign period. For example, the Portuguese team published frequent updates in a newspaper. The Croatian team published reports according to the official campaign-finance reporting cycle, and directly contacted the electoral management body when identifying a breach of the campaign finance law. Every team shared their findings at an event with relevant electoral stakeholders and CSOs.

Table 2: Comparison of SMM Approaches

Country	Pre-election data collection	Post-election data collection	Platform	Data Access Tool	Analysis Tool	#posts	#accounts	Sample selection	Area of focus	#reports published
Austria	31 days	7 days	Facebook, Twitter, YouTube	Facebook API, Facebook Ad Library, Twitter API, YouTube API	Python, R, R Shiny	26,000 posts /1.1. million comments	146	Political actors (i.e. Politicians, parties, media and user comments)	Total activity, user engagement	1
Portugal	30 days	90 days	Facebook, WhatsApp	CrowdTangle API, Public WhatsApp links	Own tool, CrowdTangle, Forsight Brandwatch	138,000 posts	85	Pages and groups not officially linked to parties or media sharing political content (above a threshold of members and likes with recent posts)	Post reach, audience growth and key narratives	Risk Assessment + 2 Reports
Poland	41 days (not including silence period)	0 days	Facebook	NewsWhip	Excel	8,848 posts	74	Political actors (i.e. Political parties, politicians, influencers/ journalists and media outlets)	66 narratives, total activity, user engagement	1
Romania	9 days	14 days	Facebook	NewsWhip	Excel	14,507 posts	102	102 public pages (Media, influencers, public officials)	117 narratives, total activity, user engagement	1
Croatia	n/a*	n/a*	Facebook	Facebook Ad Library, Facebook Ad Library Report, Google's Transparency Report	Microsoft Power BI and Tableau	n/a	11	Politicians and political parties	How Facebook political ad spends match to official reporting	1Risk Assessment + 3 Reports

*Data analysed by state financial reporting periods rather than pre- and post- election periods

** Based on the final report (includes 47 public pages and 38 public Facebook groups)

3. Lessons Learned to help future SMM teams



Preparation Phase

- **Teams require a combination of local political knowledge and some data analysis capacities**

As seen by the various organisational types and team structures, many different combinations can work. Many of the teams included (a) individuals with a political, social science or electoral observation background and (b) data scientists or (c) someone who met both of these criteria. The Polish case shows that it's possible to work with a team conducting social media monitoring for the first time with the proper training and structure in place. All teams were made up of four to seven individuals, and in most cases the number of team members corresponded to the quantity of data studied.

- **Collaboration with other organisations early on can improve the quality of your study**

Building on pre-existing knowledge (e.g. problematic accounts) or external lists of key political actors helped the teams streamline their preparation process. In the Austrian case, where this information was not available, they conducted a pre-monitoring period to acquire this information.

The Romanian team coordinated with other local actors to prevent overlap between studies. Coordination also helped some teams identify known disinformation accounts, topics or trends within their local context.

Although not tested in these cases, some groups suggested future coordination with academia could lead to AI solutions to simplify the work of CSOs. However, as tested in the Austrian case, AI solutions (sentiment analysis and topic modelling) are not always successful even when data is available. The Austrian team found that the results of their models were not qualitatively meaningful. This suggests that human expertise is still needed until more advanced tools are developed.

- **Including unofficial actors as well as official actors in your sample is key**

Some teams monitored only official actors (e.g. politicians, parties and media), which was critical to understanding those accounts' importance and role in the election. However, as identified by the Portuguese study, low-quality viral or false content may be spread by unofficial actors. As a result, it may be important also to monitor unofficial and non-mainstream accounts as possible sources of disinformation.

Identifying unofficial actors may be more difficult during the sample selection process. In the Portuguese case, they identified disinformation actors by filtering for accounts that discussed a popular political topic, corruption. Based on this, they were able to identify or non-mainstream accounts that were discussing politics and posting news articles. As a similar approach to identifying non-mainstream political accounts, teams can identify the top domains discussing the election based on a keyword search and isolate the non-mainstream accounts.

- **Starting from a non-partisan perspective is important**

To produce a non-partisan output, it is essential to select a sample that looks at both sides of any given debate. For example, instead of examining only far-right groups, it is worth understanding the full spectrum of voices to identify differences and similarities. Rather than being exclusively problem oriented, neutrality should be considered by social media monitors.

- **Teams should not expect to follow one "golden" SMM methodology**

Projects did not follow one step-by-step approach, but were instead organisational and context specific and limited by data access restrictions and team capacities. Instead of looking at the entire sea of official and

unofficial actors for post activity, political advertisements, disinformation and key narratives, teams actively decided to limit their project scope. As a result, SMM can be thought of as a series of smaller questions and analyses that make up a bigger whole.

To help teams [build their own methodology](#), DRI created a Digital Democracy Monitor toolkit to help teams come up with their own step-by-step approach based on their platform of choice and prioritised areas of focus.



Data Access, Collection and Analysis

- **Accessing Facebook data was the most difficult and YouTube was the easiest**

Access to Facebook data was critical for all teams as the most important platform in all countries. Following the Cambridge Analytica scandal, access to this data by Facebook has been increasingly difficult.

In terms of recommendations to access Facebook data, one partner noted that getting in contact with a “human representative” was the only way to eventually get access after a lengthy three-month period. Local media organisations may have CrowdTangle access, so partnerships may be beneficial for both outreach purposes and data access. When CrowdTangle or Facebook API access is not available, teams must use off-the-shelf data access platforms like NewsWhip. If strategic CrowdTangle partnerships are not possible, applying for access as an individual researcher is also possible.

The Austrian team noted that YouTube API access was the easiest, although programming skills are needed to use the resource. The platform has at-length documentation and resources to use the service, including code in different programming languages to access the API. No wait time is needed to gain access credentials.

Despite Facebook data being the most difficult to access, it is the only platform that provides free data access to researchers without requiring programming skills via an API (as with Twitter or YouTube). In this sense, once access to CrowdTangle is granted it may be the easiest tool to use for non-programmers.

- **During the data collection process, access may be unreliable**

Even once a team received data access, the APIs were unreliable as access rules would change. The Austrian team noted that [Tweets disappear after 7-10 days](#), per company policy, so they need to be constantly downloaded. Coding requirements to work with the Facebook API constantly changed without warning so the team’s code broke several times while collecting data.

- **Saving data in real time is critical in case platforms delete posts**

The teams noted challenges with social media platforms removing problematic or disinformation content from the data access platforms/APIs. As a result, it was critical for teams to save screenshots or image files of potentially problematic posts.

- **Facebook Ad Library/Report do not provide comprehensive data**

The Croatian team in particular attempted to match political advertisement data to candidate’s official candidate reporting. This proved difficult because they were not able to access specific time periods of interest, but instead buckets of dates provided by Facebook. These buckets did not correspond to the same candidate financial reporting periods. When CSOs cannot access exact social media ad expenditures, they cannot precisely match this information to officially declared expenditures in the candidates’ financial reports. Additionally, the ad identification numbers do not match the post identification available via the Facebook API, so it is not possible to track user interactions with advertisements. Additionally, the amounts paid are available in buckets (i.e. 500-4999 EUR) which makes it impossible to determine the demographics that candidates targeted with their advertising.

- **Even when monitoring for only a few weeks, you will end up with an abundance of data**

Several of the teams noted an initial feeling of concern by not starting social media monitoring earlier. However, all teams ended up with significantly more data than they were expecting and the bigger challenge was actually sifting through this abundance of information. This may be particularly dependent on your data access tool. For example, Newswhip allows researchers to extract 300 posts at one time, while CrowdTangle allows researchers to access an unlimited number of posts.

- **Non-text media is increasingly important and challenging to analyse**

Many of the teams experienced challenges with analysing non-text media including images and videos. Some teams saved screenshots or image files, although this is time intensive. Videos, in particular, are difficult not only to save but also to classify because they require a manual coder to watch several minutes of media rather than reading 280 characters. This is increasingly important given the rise of platforms like Instagram, TikTok and YouTube.



Publication and Outreach

- **Events with electoral stakeholders, media and government officials increase the project's impact**

All teams presented their reports at events, which helped deliver their key findings to decision-makers and increase their reach to relevant stakeholders.

- **Timely output is exhausting but may have real-time impacts on decision-makers**

This is especially seen in the Croatian case, where Gong's public notification of reporting errors may have helped contribute to a correction by the candidate in question during the next financial reporting period.

Cooperation with news organisations notifies citizens in real time

The Portuguese team collaborated with a local news organisation, which helped them promptly inform voters about identified disinformation narratives to be aware of online.

- **Further report-writing resources are helpful for maximising impact**

How to publish a project's results for maximum impact is important to consider at the beginning of a project. As this is not always an easy question to answer, DRI has developed [some tools](#) which are available in our Digital Democracy Monitor toolkit.

Teams must first determine what they are trying to achieve with their findings depending on their organisational mission and project goals. For example, publishing a two-page summary with key findings may help organisations speak to policymakers. A 20-page report is nonetheless important, but may speak to more technical audiences looking for an in-depth analysis. An interactive tool as seen in the Austrian case may be a more user-friendly way to display a project's results. Teams must consider their relevant audiences and how to best reach these stakeholders via their writing style and page length.

For future research and comparable results across countries, developing further templates may also be beneficial to social media monitoring teams.

4. Recommendations to help future SMM efforts

Based on the lessons learned, future social media monitoring efforts could benefit from action by specific stakeholders:



Government

- **Require companies to provide reasonable access to CSOs when monitoring a public space**

Social media companies are not providing the needed data sources, and self-reporting methods do not include sufficient information. See below recommendations more some more specific data needs.

- **Invest in more social media monitoring projects at the local level**

The learning curve for SMM is steep, but once a CSO conducts one or more studies, they have the long-term infrastructure (e.g. data access, knowledge) to produce more reports. There is a growing need to establish more SMM teams, especially given the recent Covid-19 “info-demic”. CSOs need the proper resources to develop their reporting capacities and provide information in real time.



Social Media Companies

- **Provide reasonable access to CSOs when monitoring a public space.**

For Facebook in particular, many barriers make it difficult for researchers to access the available data. Also, this process is very selective and unclear, with unexplained rules. One team received access to CrowdTangle while another received only limited data access via the Facebook API. Tech companies should provide a direct contact channel that electoral observers and civil society organisations can access to ask for data for research.

- **Provide access to deleted data for researchers**

Social media platforms are deleting problematic posts, which is critical data for election social media monitoring teams to study. When this data is not available, it is not possible to understand its potential impact on an election. As one possible solution proposed by the Portuguese team, creating an API from a mirrored social media platform with all posts archived would allow researchers to gather these deleted posts. Further research is needed to understand if this issue conflicts with the EU’s “right to be forgotten” legislation.

- **Provide access to additional aggregate metrics related to user engagement (e.g. clicks)**

This is non-private data, but the information would help researchers know how many people actually clicked a link to read beyond the headline. This metric indicates potential impacts of a post beyond just a like.

- **Improve data access stability for researchers**

Unstable data access creates major hurdles for social media monitoring teams. During elections this matters because losing a day of data may prove to be a significant shortcoming.

- **Increase the quantity data that researchers are permitted to download**

Currently, it is difficult for teams to download the required volume of data for research in a given attempt which

causes additional challenges. In the case of the Facebook API (not CrowdTangle), the amount of data permitted depends on the number of users a Facebook App has, which is problematic when research does not involve designing an App (instead, a server-to-server App). In the case of Twitter, free data access is limited to a certain number of days and Tweets.

- **Expand the Facebook Ad Library in order to hold candidates accountable during elections**

Researchers need access to specific time periods rather than ranges or data to do so. They also need to understand data differentiation between paid ads and boosted content as both have an impact on citizens. For further details see above section (p. 14) and DRI's [recommendations](#).



CSOs and Universities

- **Make automated language processing solutions accessible to CSOs**

CSOs could benefit greatly from AI tools to detect disinformation, hate speech or parse through large datasets more quickly.

- **Publish a guide or tool to ethically use, save and publish social media data for research**

A number of teams expressed uncertainty on how to proceed regarding data ethics and would have benefited from such a guide. As a start, this may entail a simple guide with ethical guidelines related to handling social media data in a GDPR-compliant manner. Additionally, building a platform to save social media data in a GDPR-compliant manner would help teams in practice. Furthermore, a platform for the gathering of open source tools that could be used for monitoring purposes would also benefit social media monitoring teams.

- **Build a repository of sample studies to help organisations get started**

A further repository of sample studies may help new SMM organisations match their organisational mandate to a methodological approach. For example, it would be helpful for organisations to see all other studies conducted in their country or for a gender-focused organisation to see sample studies on gender-based bias or harassment online.

- **Train or collaborate with journalists to conduct this type of work**

Involving journalists in SMM may help deliver findings to a broader audience including citizens directly. Journalists should not only debunk individual stories but look deeper into how these stories were spread on a systemic level.

