Supporting policy-making with social media and e-participation platforms data

Simonofski, Anthony; Fink, Jerome; Burnay, Corentin

Published in:
Government Information Quarterly

DOI:
10.1016/j.giq.2021.101590

Publication date:
2021

Document Version
Peer reviewed version

Supporting Policy-Making with Social Media and e-Participation Platforms Data: a Policy Analytics Framework

Anthony Simonofski[ab], Jerôme Fink[a], Corentin Burnay[a]

[a] Namur Digital Institute, University of Namur, Rue de Bruxelles 61, 5000 Namur, Belgium

[b] Faculty of Economics and Business, KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium

Abstract

E-participation enables citizens to have an impact on policy-making through electronic means. Two of the most popular channels are social media and dedicated e-participation platforms. However, the ideas, comments, discussions of citizens on these two channels generate a lot of data to be processed by political representatives or public agents afterwards. Despite the existence of various techniques for social media analytics, literature is scarce regarding the analysis techniques to mine e-participation platforms as well as the possible combination of insights between the two channels.

In order to address these gaps, we design a policy analytics framework to leverage insights from e-participation platforms and social media through relevant data analytics to support policy-making. In order to do so, we rely on the Design Science Research methodology. Through the analysis of four different cities in Belgium (Liège, Mons, Marche-en-Famenne, Leuven), we identify policy-makers' requirements and needs of information from platforms and social media. Then, we explore data analysis techniques to address those requirements. Finally, we design an actionable framework, present it as an interactive dashboard and iteratively test it on the case of Liège. This policy analytics framework supports each step of the traditional policy-making process with appropriate data analytics applied to the two sources.

Keywords: E-Participation, Social Media, Policy Analytics, Design Science Research, Dashboard

1. Introduction

The impact of citizens on the decisions taken by political representatives and on policy-making in general is labeled as "citizen participation" (Arnstein, 1969) and is not new. However, this participation can be further stimulated through the use of Information and Communication Technologies (ICT), making it more accessible but also cost-efficient. The use of ICT to support participation is labeled as "e-participation" and defined as "the use of information and communication technologies to broaden and deepen political participation by enabling citizens to connect with one another and with their elected representatives" (Macintosh & Whyte, 2008).

e-participation encompasses different participation channels (Berntzen & Johannessen, 2016; Simonofski, Snoeck, Vanderose, Crompvoets, & Habra, 2017). These channels can be structured into two main categories: government-led initiatives, and citizen-led initiatives (Lee & Kim, 2014; Porwol, Ojo, & Breslin, 2016b). Government-led initiatives are for example e-voting, online surveys or e-participation platforms, whereas citizen-led initiatives could be blogs or forums where citizens can write their ideas and react to ideas of fellow citizens (Lee & Kim, 2014). Porwol et al., (2016b) mention that the most common citizen-led initiatives are discussions on social media platforms such as Facebook or Twitter, since this is the area of the Internet where citizens are most active in general. In this paper, we will focus on two of these channels: e-participation platforms (Government-led) and social media (Citizen-led).

Macintosh et al. (2009) underline the importance of building synergies between these two channels. Indeed, e-participation platforms are often built or bought to stimulate participation among the population in a structured way but often lack engagement (Macintosh et al., 2009; Toots, 2019). On the other hand, citizens tend to engage quite easily on social media with spontaneous political discussions but in a less structured manner (Porwol et al., 2016b), which makes it harder to exploit for policy-makers. Furthermore, it is particularly interesting for policy-makers to gather insights from both channels as the two channels are not used by the same groups of citizen. On the one hand, social media gather numerous citizens sharing ideas without any real political intent. This channel has therefore the potential to be very inclusive. On the other hand, e-participation platforms gather more motivated individuals who are already politically active (Berntzen & Johannessen, 2016). On top of those respective limitations, there is a gap around the data analysis techniques that could be applied to make a better use of the data available on these two channels, especially for online platforms. The available information can be overwhelming and difficult to process for policy-makers, who would therefore benefit from a better process/system to extract meaningful observations or even recommendations. Several data analysis techniques have been mentioned in the literature such as Marianne et al. (2019) for e-participation platforms or Belkahla Driss, Mellouli, & Trabelsi (2019) for social media. Even though previous research focused on data analysis to improve policies (e.g. Policy modelling (Gilbert, Ahrweiler, Barbrook-Johnson, Narasimhan, & Wilkinson, 2018), Policy analysis (Kolkman, 2020)), the possible combination of data analysis from these two specific channels and embeddedness in a broader policy-making process are two questions that remain to be explored. As an answer, this paper suggests an actionable policy analytics framework for policy-makers to harness the ideas, discussions, and feedback provided by citizens. This framework supports each step of the policy-making process with appropriate data analytics to harness insights from the two channels.

This paper is structured as follows. In Section 2, we discuss the relevant literature about data analysis techniques for e-participation platforms and social media. Furthermore, we explain how these techniques can impact policy-making. In Section 3, we explain the different methodological stages of Design Science Research we followed to build the framework as well as the four selected cases to support this process. In Section 4, we detail the requirements of practitioners regarding the framework, we describe the framework, and we test it on the case of Liège (Belgium). In Section 5, we describe the implications for research and of practice of this paper. We also detail its limitations and introduce further research leads. Finally, in Section 6, we summarize the findings of our research.

2. Background

2.1. Data Analytics for Policy-Making: Policy Analytics

A public policy is established through a policy-making process: a type of decision process with specific characteristics (Tsoukias, Montibeller, Lucertini, & Belton, 2013). Many theoretical frameworks exist to describe the chronological steps this policy-making process consists of. In this paper, we rely on the broadly accepted work of Howlett, Ramesh, & Perl (2009) that describes the Public Policy-Making Process into five stages:

1. Agenda setting: framing of the problem and exploring the needs for a policy;
2. Formulation: developing policy options and informing them;
3. Decision: selection of the option among a range of options;
4. Implementation: enactment of the selected policies through legislation, regulation, planning;
5. Monitoring: evaluation of the effects of the past policies.

A public decision process is conducted through consultation, participation, and mutual consent of a large number of diverse actors: different organizations, citizens and other individuals, interest groups, and officials (De Marchi, Lucertini, & Tsoukiàs, 2016; Kim, Trimi, & Chung, 2014). Therefore, the decision-making process of a government usually takes much longer than in a business (Kim et al., 2014). As already discussed, e-participation enables this participation and consists of different channels to collect insights from the citizens and thus to make the consultation process more efficient. However, the collection of insights, their analysis, their impact on decision-making need to be managed by key stakeholders. Indeed, a poor management of the e-participation process and integration of its insights into policy-making can lead to negative outcomes (Lee & Kim, 2014).

In order to understand the impact of data analytics on policy-making in the context of e-participation, we choose to rely on the recent policy analytics literature, defined as the use of data and analytical techniques to support policy decisions (Gil-Garcia, Pardo, & Luna-Reyes, 2018). Janssen & Helbig (2018) studied how the policy-making process is changing due to new evolutions in governance and technologies. They underline the need to analyze the large amount of content that is produced, among other sources, on e-participation channels. A lot of recent studies labeled this process differently with terms such as policy modeling (Gil-Garcia et al., 2018; Hamza & Mellouli, 2018), policy informatics (Cronemberger & Gil-Garcia, 2020; Gil-Garcia et al., 2018), policy analysis (Kolkman, 2020) (Hamza & Mellouli, 2018), data-based policy (Nam, 2020), government data analytics (Cronemberger & Gil-Garcia, 2020), and public policy analytics (Tsoukias et al., 2013).

Daniell, Morton, & Ríos Insua (2016) also underline the specifics of using data analytic for public decision-making compared to decisions taken in the private sector. One such specificity is the influence of citizen participation on the decision-making of policy-makers and the need for them to balance several sources of information to design policies. Cronemberger & Gil-Garcia (2020) suggest that policy analytics is in fact a data management process that consists of conceptualization of the problem, data collection and preparation, and data analysis and visualization. They underline that the conceptualization of the problem and the consequent information need more research. Matheus, Janssen, & Maheshwari (2020) discusses dashboards as a way to bundle the data analysis techniques and present it to policy-makers. The design principles reported in this paper will also fuel our research. These studies show that the theory-building of policy analytics is currently ongoing. However, none of these studies have studied it in the context of e-participation. Therefore, what policy-makers expect from analytics to study e-participation data remains unexplored. This is in line with the global perspective on Policy Analytics advocated in the above-mentioned papers and currently understudied.

In the next subsections, we detail what we mean by data analysis techniques and how the techniques are currently used for e-participation platforms and social media.

2.2. E-Participation Platforms Analytics

e-Participation platforms are online platforms that facilitate the implementation of e-participation. Different features are offered on these platforms (Schossböck, Sachs, & Leitner, 2016) and authors suggested various maturity models to explain them. For instance, Abu-shanab (2013) highlighted five levels of e-participation: e-Information, e-Consultation, e-Involving, e-Collaboration, and e-Empowering. E-informing consists in providing details about relevant policies to citizens. E-consulting is a two-way communication channel where stakeholders and citizens can discuss their opinions and contribute to the study of issues. E-Involving ensures that the opinions of citizens concerned by certain issues are assessed. E-collaborating is a two-way channel where stakeholders and citizens are partners in generating solutions. E-empowering gives citizens the power to make decisions and monitor policies. Different platforms will focus on different requirements, among the five previous categories. In summary, e-participation platforms enable communication between the stakeholders in decision-making, voting, or debating processes (Berntzen & Johannessen, 2016).

Research devoted to the analysis of data produced on e-participation platforms remains scarce, despite the growing number of ideas submitted on these platforms. In most cases, the analytics possibilities to interact with ideas found on the platforms come down to some basic drill-down and filtering features, for instance to classify ideas according to some overall themes or to divide them by some user-defined socio-demographic characteristics (Fedotova, Teixeira, & Alvelos, 2012). Often, public agents and political representatives consider the manual processing of the data as satisfactory when the amount of data is manageable, despite huge time and efforts dedicated to the analysis (Panopoulou, Tambouris, & Tarabanis, 2010; Rose & Saebo, 2010). The research related to the analysis of data from these platforms to impact the whole policy-making process remains scarce.

Indeed, a vast body of work have focused on the use of analytics (such as topic modelling, or visualizations) to harness insights from e-petitioning platforms (Hagen, 2018; Hagen, Keller, Yerden, & Luna-Reyes, 2019; Hale, John, Margetts, & Yasseri, 2014). However, harnessing insights from e-participation for all steps of the policy-making process is less explored as e-petitions focus on identifying themes for agenda-setting but less on the discussions and alternatives suggested by citizens. In a recent study about the collection of citizens' input for policy formulation Marianne et al. (2019) defined a thematic visualization (taking the form of a tree) of citizens' ideas based on the exploration of the platforms of Liège and Mons in Belgium, thereby suggesting more advanced analytical features could also generate interest and added-value for policy makers.

2.3. Social Media Analytics

A major limitation of the previously discussed platforms resides in the lack of integration with citizen-led participation that can be found on social media (Porwol, Ojo, & Breslin, 2016a). Some systems now integrate an interface with social media platforms and tend to receive researchers' preference (Dolson & Young, 2012). Social media can be considered as a public sphere where discussions and political debates can take place (Clay, 2011). It can also be considered as an effective mean for citizen participation. For instance, Sandoval-Almazán et al. (2017) list 17 principles to foster citizen engagement on social media. Furthermore, in a recent empirical study, Bonsón, Perea, & Bednárová (2019) conclude that Twitter can lead to increased citizen engagement depending on the media and content of the tweets of the local authorities.

Several data analysis techniques exist to exploit data generated by social media. For instance, Grover, Kar, Dwivedi, & Janssen (2019) studied how social media discussions have the ability to impact the outcome of national elections in the United States. However, in the context of this study, we focus on how social media analytics can support policy-makers' decisions. According to Porwol et al. (2016b), there is a need for infrastructure to enable decision-makers in accessing and extracting relevant information about ongoing citizens' discussions on social media platforms.

Authors have made attempts to facilitate this data analysis. Belkahla Driss, Mellouli, & Trabelsi (2019) suggest a framework to exploit social media data from Facebook by policy-makers through natural language processing. (Androutsopoulou, Charalabidis, & Loukis, 2015) focuses on the use of social media analytics to monitor public policies. Rodríguez Bolívar (2018) explores the use of social media analytics in the context of civic engagement and concludes that social media is more used as a unidirectional communication channel by governments. Maragoudakis, Loukis, & Charalabidis (2011) perform a review of opinion mining methods to support e-participation and identify three levels of analysis for documents, sentences, and features. None of the previously cited research focused on the combination of these insights with the ones from e-participation platforms. Once again, this gap is mitigated by studies combining insights from social media and e-petition platforms. For instance, (Asher, Leston Bandeira, & Spaiser, 2017; Cihon, Yasseri, Hale, & Margetts,

2016) use data from an e-petition platform and Twitter. However, they compared the engagement on both platforms as well as the process in itself. In this study, we aim to combine insights from both platforms to support the policy-making process.

3. Research Approach

3.1. Research Questions

First of all, several data analysis techniques have been mentioned in the Background to exploit data from both platforms and social media. However, we take the viewpoint of the policy-makers and start from their need for information coming from both channels. This will serve as initial requirements to develop our framework. Therefore, we formulate this first Research Question (RQ):

- RQ1: Which information do policy-makers find relevant to extract from social media and from e-participation platforms to support the policy-making process?

Second, after establishing the information needs of policy-makers, the different data analysis techniques need to be explored. Information can be found in the literature about these techniques but the combination of relevant techniques to harness data from these two sources is not tackled. Furthermore, there is a lack of advanced reported techniques for e-participation platforms. This brings us to the definition of a second RQ:

- RQ2: How to extract and combine valuable insights for policy-makers from social media and e-participation platforms?

Finally, after exploiting the data, we need to study how it inserts itself into a more holistic process followed by the policy-makers and supports their daily work. Therefore, we formulate our third and last RQ as follows:

- RQ3: How can data analytics, applied to social media and e-participation platforms, support the policy-making process?

In order to design and structure our research model, we relied on the integrated e-participation analytics process of Porwol et al. (2016b) and on the policy analytics process of Cronemberger & Gil-Garcia (2020). However, in this paper, we dig deeper into the data-processing aspect of these two models. Figure 1 positions our research questions visually on our research model.

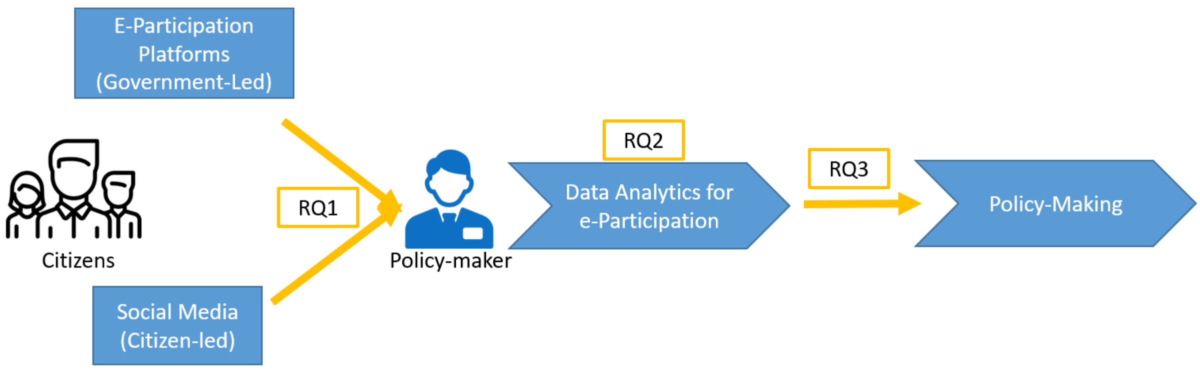

Figure 1. Research Questions

3.2. Methodology

3.2.1. Design Science Research Process

To develop the actionable policy analytics framework, we apply the Design Science Research (DSR) methodology. This methodology is based on the methodologies suggested by (Hevner, March, & Park, 2004; Smith & March, 1995) as it aims at creating an output artifact linked with technology to serve human purpose. Furthermore, this methodology has been applied in previous studies to develop tools to support policy-makers in managing e-participation (Porwol et al., 2016b; Simonofski, Asensio, De Smedt, & Snoeck, 2019). DSR consists of three research cycles to build the artifact: Relevance Cycle, Rigor Cycle and Design Cycle. To ensure the Relevance Cycle of DSR, we aim at identifying the objectives of the solution to tackle the problem identified in the practical environment (RQ1). To ensure the Rigor Cycle of DSR, we identified and motivated the problem to be tackled through an exploration of the literature to ensure a contribution to the theoretical knowledge base (RQ2). Third, in order to ensure the Design Cycle of DSR, we will develop, demonstrate and evaluate our artifact: the framework (RQ3). Each researcher was assigned as "leader" in a specific cycle depending on his requirements engineering (RQ1), data analytics (RQ2) and dashboarding (RQ3) backgrounds. First, to address RQ1, we studied four different cases (Liège, Mons, Leuven and Marche-en-Famenne) through interviews to ensure the relevance cycle of design science research and identify the objectives of our artifact. These four major Belgian cities are further described in Section 3.2.2. The objective of this cycle is to better understand the participation channels the cities are using, which information they are seeking on each channel, which data analysis techniques they are using, and what would be their requirements for our framework. In order to analyze the interviews, we applied thematic content analysis (Anderson, 2007). This method enables to link similar themes from every interview to each other, making it easier to analyze what is being said and how it compares with the other interviews. We categorize the results from our interviews (textual data) to concrete requirements to drive the development of the framework.

Second, to address RQ2, we perform a gap analysis based on existing research and existing tools. This allows to ensure the rigor cycle of Design Science Research to ensure that the combination of techniques used in our framework will indeed contribute to the theoretical knowledge. This gap analysis was performed through a literature review (reported in Section 2). An exploration of the available data in the four cases was conducted to identify the relevant data analysis methods. Then, we investigated techniques for clustering documents similarity and sentiment analysis presented in (Bengfort, Bilbro, & Ojeda, 2018) to find methods suited to our requirements. We retained the methods for which no text annotation was needed and leading to satisfactory results. The approaches identified for text clustering were the Latent Dirichlet Allocation and Distance Metrics Clustering. For sentiment analysis we identified the polarity techniques. The retained techniques will be detailed in Section 4.2. The curated data and the exploration code are publicly available in a GitHub repository[1]

Third, we develop and evaluate our framework (the design artifact) through several iterative design cycles to address RQ3. As explained in (Smith & March, 1995), the development of the framework is iterative in nature. Based on the underlying constructs identified in the literature review, the requirements identified in four cases, and the technical possibilities explored, we propose a mapping of data analytic techniques to the relevant stages of the policy-making cycle to address the elicited requirements. This mapping of techniques was then bundled and presented as an online dashboard. The dashboard was implemented using Power-BI, the dashboard solution of Microsoft. The tool is very simple and intuitive to use for non-tech users, offers advanced analytical features in line with the requirements of policy makers, and is leader on the data analytics market. We use this Power-BI dashboard as a visual aid to iteratively improve and validate our framework through follow-up think-aloud interviews and through its application to the use case of Liège. Indeed, Design Science Research argues for the realization of an artifact in its environment to demonstrate its feasibility and effectiveness.

3.2.2. Description of the cases

The four cases were selected based on several factors. First, the cities all had an e-participation platform (e-collaborating platforms where citizens can discuss and share concrete solutions) and an the active presence on social media. All cities were at the same stage of maturity regarding e-participation and social media which decreased the risk of contextual factor influence. Second, the interviewees were available. No specific ties were previously made with the four cases but direct contact was easily feasible at each step of the research. Third, the cities are all located in Belgium. Indeed, according to a recent barometer (Desdemoustier & Crutzen, 2017) and previous studies (Simonofski, Asensio, et al., 2019; Simonofski, Vallé, Serral, & Wautelet, 2019), the policy-makers in Belgium tend to engage in participation more and more in the context of smart city development.

[1] Anonymized link due to double blind review

Mons is a Walloon city of 95,299 inhabitants that positions itself as a creative city with a living lab, mostly through collaborations with the university of Mons. They invested in an e-participation platform (https://mons.citizenlab.co/fr-BE/) to collect ideas in all domains from citizens. Furthermore, they have a strong presence on Facebook. Leuven is a Flemish city of 101, 396 inhabitants. Recently, left-wing political representatives positioned it as a participatory city and invested in an e-participation platform (https://leuvenmaakhetmee.be/) to explore citizens' ideas at large scale but also for more focused projects (e.g. mobility). Leuven has the specific to heavily link its platform to social media (Twitter and Facebook) by sharing the ideas of the platform to generate discussion and visibility. Marche-en-Famenne is a smaller city of 56,733 inhabitants and positioned itself as one of the first smart cities in Wallonia. Therefore, they invested in an e-participation platform (https://jeparticipe.marche.be/) used for targeted citizen consultation and participatory budgeting. Furthermore, they have a strong presence on social media (Facebook, Twitter, and Instagram).

Liège, one of the largest cities in Belgium (195,965 inhabitants) and second largest in Wallonia, engages in participatory actions to develop the public policies and can be considered as having the most experience out of the four cases. Since 2017, Liège worked with an e-participation platform called "Reinventing Liège" (https://www.reinventonsliege.be/) and adopted a new platform in 2019 called "Liège 2025" (https://www.liege2025.be/). However, Liège faced challenges in the processing of the data as more than 1000 projects and ideas were reported on the platforms. In terms of social media presence, Liège has a Twitter account[2] and a Facebook page but hardly analyzes the insights from these channels. This case was chosen, not only to elicit the requirements in RQ1 but also as main data source for RQ2 (as the data from both channels was easily accessible) and as a validation case for RQ3 (as the practitioners were open for collaboration and close interaction to give context to the data and feedback about the research). Regarding the data sources in RQ2, Facebook data were discarded as political posts are more often posted on Twitter than Facebook in Liège and Facebook also tends to discourage the analysis of posts and comments by making them more difficult to retrieve automatically. The technical exploration was performed on the 988 ideas collected by the city officials via the platform (available for re-use on their Open Data portal[3]) and more than 3000 tweets containing the word "Liège" collected between April and May 2020. First, the exploration focused on the quality of the dataset by performing simple statistical analysis such as the number of Twitter accounts involved, the mean number of tweets per account, the most frequent word, the most active Twitter account. These observations allowed us to ensure that all the tweets in our dataset are relevant to our case. For instance, we noticed that tweets containing the German word *liege* were collected. Therefore, we filtered the tweets in order to only keep the ones written in French. We also filtered tweets with less

[2] https://twitter.com/VilledeLiege
[3] https://opendata.liege.be/explore/dataset/reinventons-liege/information/?disjunctive.categorie

than 100 characters as we do not expect short tweets to convey valuable information for decision makes. At the end of this process, 1962 tweets relevant to our case were retained. Basic pre-processing was also conducted on the text data such as removing stop-words and smiley. The texts were also stemmed to reduce the vocabulary size of the text documents and ease the topic modeling process. Thereafter, we therefore decided to apply the framework, step-by-step, to the case of Liège to illustrate its relevancy. Furthermore, the two interviewees from Liège were available for follow-up interviews about our research.

Table 1 summarizes the information about interviewees for the four cases. On average, the interviews lasted an hour and were semi-structured around an interview guide (Drever, 1995). Depending on the digital and participation maturity of the interviewed city, the questions sometimes focused more on one specific channel in-depth with additional probing questions. The complete interview guide can be found in Appendix 1. We performed interviews until we reached code saturation, meaning that no new codes (and thus requirements) were identified after five interviews (Guest, Bunce, & Johnson, 2006). The three following were still conducted to ensure the completeness of the requirements and help us quantify their importance, i.e., are they frequently or less frequently mentioned by policy-makers? The number of interviewees per case is relatively small as the core teams in charge of participation in each of the four cases are small as well. However, the interviewees had different responsibilities in the cities, allowing to give complementary information about e-participation platforms, social media, and policy-making.

Table 1. Interviewees for the four cases

#	RQ	Case	Function	Relevancy
1	RQ1 RQ2 RQ3	Liège	Strategic Planning Coordinator and Smart City Manager	Manages the participation activities of the city (operational level)
2	RQ1 RQ2 RQ3	Liège	Head of Strategy and Planning Department	Coordinates the participation activities of the city (strategic level)
3	RQ1	Mons	Researcher	Supports the city’s participation activities
4	RQ1	Mons	Political Representative in charge of participation	Coordinates the participation activities of the city (strategic level)
5	RQ1	Leuven	Head of urban development department	Coordinates urban planning activities, with a particular focus on participation
6	RQ1	Leuven	Communication Manager	Handles the internal communication related to the participation platform and the external communication on social media

7	RQ1	Marche-en-Famenne	Communication Manager	Handles the internal communication related to the participation platform and the external communication on social media
8	RQ1	Marche-en-Famenne	Political Representative in charge of participation	Coordinates the participation activities of the city (strategic level)

4. Results

4.1. Identified Requirements for Policy Analytics

In this section, we present the stakeholders' requirements about the extraction and analysis of data from the two channels to support the policy-making process. Table 2 summarizes the requirements extracted from the interviews, the stages of the policy-making cycle they are linked to, the source of data (social media or platform), and the occurrence of the requirements across the interviews.

Table 2. Summary of the requirements

N°	Requirement	Stage	Source	Occurrence
R1	Identify recurring tendencies from the population to put in the political agenda	Agenda-Setting	P / SM	8
R2	Check the match between the themes discussed on social media and on platforms to refine big tendencies into focused projects, alternatives, and actions	Formulation	P / SM	7
R3	Extract the opinion of public agents about feasibility of the projects qualitatively	Formulation	P	3
R4	Extract the opinion of citizens qualitatively (e.g. comments, link between projects, implementation ideas)	Formulation / Decision	P / SM	6
R5	Extract the opinion of citizens quantitatively (e.g. voting, number of ideas, socio-demographics)	Decision	P	5
R6	Support the decisions of policy-makers visually	Decision / Transversal	P / SM	7
R7	Use the wording of the citizens when formulating the policies	Implementation	P / SM	1

R8	Monitor citizens' opinion on social media about the evolution of the actions taken by policy-makers	Monitoring	SM	7
R9	Identify hot topics that stimulate discussion and debate (emotions, feelings), especially in reaction to the city's actions	Monitoring / Agenda-Setting	P / SM	3
R10	Communicate and vulgarize the insights to non-expert citizens to receive feedback	Monitoring	P / SM	3

To complete the insights summarized in the table, we can mention that the Agenda-Setting, Formulation and Monitoring stages were the main focus of the interviews. Indeed, in the Decision stage, the participation of citizens was mentioned by the interviewees as only one criterion among other, more important, criteria (such as feasibility, political agenda, coherence with other actions of the city, etc.). However, the visual support for decisions was a transversal discussion throughout the requirements. In the Implementation stage, citizen participation was reported either as less important or more often done offline through co-creation workshops, assemblies, etc.

Furthermore, other requirements were discussed in the interviews but fall out of the scope of this paper: requirements about the data management process (e.g. cleansing of the data) and requirements about complementarity of offline and online participation (e.g. how can the data extracted from these channels support offline activities).

4.2. Policy Analytics Framework: Exploration of Data Analysis Techniques

The policy analytics framework is structured around the five steps of policy-making as the requirements were easily mapped to each of the step. The framework consists in a sequence of techniques (or combination of techniques) mapped to the relevant stages in order to address the requirements. Not all requirements were addressed as we focused on the most important requirements in terms of occurrence. Furthermore, not all requirements could be addressed in a single framework to keep a balance between completeness and usability by practitioners. Therefore, requirements 3, and 10 were left out of scope for this paper as they had the lowest occurrences. Requirement 8 was still taken into account as the same technique was used for the requirement 9 and Requirement 7 is taken into account in a preliminary manner. **Error! Reference source not found.** summarizes the elements of the framework in a visual way.

Requirement 6 **(R6)** is here considered as a transversal requirement to provide the decision support to practitioners in a visual and integrated way. The design of such appropriate visualization for policy-makers remains an open question. Our framework attempts to account for the risks related to decision-making in an information-rich environment, and notably the risk of informational overloads and mis-shaped information structures, as discussed in (Lurie, 2004). We therefore develop a dashboard mock-up following the dashboard design insights from (Matheus et al., 2020; Sarikaya, Correll, Bartram, Tory, & Fisher, 2019) This dashboard, in order to fully facilitate the decision-making, bundles the techniques of the four studied steps in three pages. It is publicly accessible (using the following URL: https://bit.ly/2MaqC3C) and is further described and used in the validation process discussed in Section 4.3.

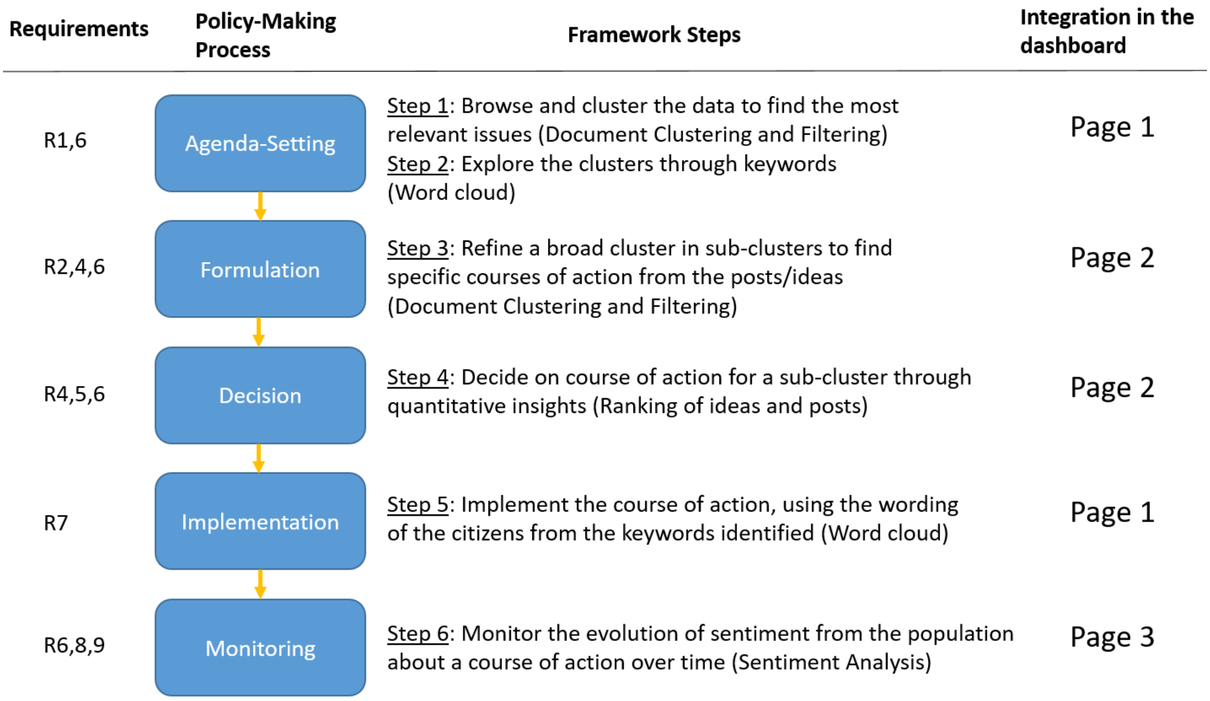

Figure 2. Framework Description

4.2.1. Agenda Setting

The decision-makers need to explore the data at their disposal in order to identify the most relevant issues and trends reported by the citizens (**R1**). At this point they have to browse all the data available to have a clear view of the situation. The gathered data consist mainly of raw text. Even if some other, more quantitative, features can be observed (number of likes, number of replies, number of posts over a period of time, etc.) none of them capture the information contained in the text. Automating the analysis of raw text data is challenging due to the inherent ambiguity of natural languages (Chowdhury, 2003). Recently, the progress in Natural Language Processing led to the creation of several tools facilitating the creation of application relying on raw text data (Porwol et al. 2016a).

For this step, we investigate methods allowing a user to explore a large volume of text data. Two categories of methods are helpful: *Document Clustering* and *Topic Modeling*. Document clustering methods aim to organize text documents into groups that can be visualized and browsed easily. The similarity of two documents are computed based on metrics such as Bag of Word (BoW) which counts the number of occurrences of each word in the document. Documents using the same words are more likely to be grouped together. The quality of the clustering depends on the quality of the methods used to extract features as shown by Liu,

Liu, Chen, & Ma (2003). On the other hand, Topic Modeling methods focus on finding the best keywords to summarize a document. Those keywords could be used later in order to filter them. The most popular methods for topic modeling are the Latent Dirichlet Allocation (LDA). Visualization libraries were also developed in order to browse the result of a LDA such as (Sievert and Shirley, 2015). Topic Modeling and Document Clustering could be used together in order to explore a dataset as they reflect different aspects of text. Another issue related to social media data, especially Twitter, is the short length of the messages collected on these platforms. In some case this could lead to poor performance when applying LDA for topic modeling. However, Hong and Davison (2010) suggest solutions to mitigate those issues by training the LDA model on aggregated tweets. The aggregation could be made by authors if we made the assumptions that Twitter users always tweet about the same subject, or by grouping the tweets containing the same hashtags.

In this paper, we propose an interface using document metrics clustering to allow the decision-makers to visualize the data collected. We use a TF-IDF encoding to represent each document and the K-means algorithm to create the clusters as proposed by Pantel & Lin (2002). To visualize the clusters, we use the t-sne algorithm proposed by Van Der Maaten & Hinton (2008). By exploring the visualization, the decisions-makers will be able to identify the different topics discussed on social media and participation platforms and to associate keywords to these topical clusters. We discard LDA for these experiments as the probability matrix created by the method cannot be considered as a distance metrics. It means that, if we plot the matrix generated by LDA for each document, we have no guaranties that document close to each other in the visualization share common properties. Its result are thus hard to visualize and explore for the user. Indeed, the point of the analytics of the agenda-setting stage is not to provide the best topic modelling but to help the user to quickly see which documents are relevant for him and which could be discarded. Therefore, we chose the most intelligible solution in this exploration.

Furthermore, by changing the number of clusters to discover, the user could find the optimal classifications for his/her case. To help the user to choose the best number of clusters, we provide measurement of a quality score for each number of clusters. An elbow method could be applied to find the optimal number of clusters. A classical quality metric for K-mean is the distortion factor which is the sum of the distances between each document and the center of their cluster. Other quality metrics for document classification exists but they often require human validation and are not fully automatics.(Hagen, 2018)(Hagen, 2018)(Hagen, 2018) A word cloud is also associated to each cluster to help the user to determine, at first sight, whether the cluster may be interesting. The user can also filter the irrelevant clusters from the visualization and only keep the interesting one for the next stages. At this point, only text data are used and we do not consider the other information that may be contained into the data set such as likes, comments, date of publication, etc. This is address in the next stage.

4.2.2. Formulation

Decision-makers focus on one of the issues identified during the Agenda-Setting to refine a broad topic into more specific ideas (**R2**). Documents are filtered using the keywords identified during the agenda-setting step to facilitate the browsing in the clusters. At this point the number of documents will be drastically reduced and the decision-maker could directly observe the remaining documents. The same Document clustering method could be used on the remaining documents in order to visualize the different points of view (or sub-clusters) about the topic. At this step we also introduce the other metadata available such as the number of reactions and comments to each social media post or citizen idea in order to highlight the ones generating a lot of discussions or reactions (**R4**).

4.2.3. Decision

Thanks to the techniques used in the two preceding phases, the decision-makers have a good understanding of the overall content present in social media and participation platforms. This content goes from overarching recurring topics (the main clusters identified in the agenda-setting phase) to the more specific sub-themes (sub-clusters in the formulation phase). Therefore, they now have to decide which course of action to follow. In that regard, the qualitative insights (**R4**) and quantitative insights (retweets, likes, dislikes, etc.) (**R5**) can help them knowing the preference of the population.

4.2.4. Implementation

We decided to leave the Implementation stage out of the scope of this framework. However, as a preliminary recommendation, we advise the policy-makers to use the wording of the citizens (identified through the keywords of the first two phases) when phrasing their policies and the communication surrounding them (**R7**).

4.2.5. Monitoring

The reaction of citizens about the implemented policies should be monitored. The most convenient way to do that is to monitor social media, because the citizens are more inclined to use the platforms to give feedback on implemented policies. The keywords found during the agenda-setting phase could once again be reused to filter the post on social media concerning the issue to monitor. We could leverage *sentiment analysis* methods on the collected posts to identify if the subject is seen positively or negatively by the citizens (**R9**). Also, by showing a timeline of the tweet activity on the subject, we could understand the reaction of the people before the implementation of the policy and compare them to the one after the implementation (**R8**). Various methods for extracting sentiments were developed especially for social media data as explained by Zimbra, Abbasi, Zeng, & Chen (2018). However, in our case, we are working with French text. The sentiment analysis methods analyzed by Zimbra are designed to work on English sentences only. Thus, we chose textblob_fr, a French extension for a popular text analysis library which provide an API for sentiment analysis. The library uses a naive polarity method counting the positive and negative words appearing in each document.

4.3. Dashboard Implementation to the case of Liège

As explained in the methodology, the mapping of techniques from the framework is presented as a dashboard for validation to practitioners. This technical solution translates data produced by the python script of our framework (described in 4.2) in a visual and interactive manner for policy-makers. The first page of the dashboard addresses the requirements for the **Agenda-Setting** phase, as shown in Figure 3. This page allows selecting the source to analyze (A) and exploring the clusters at a high level visually (B). The details about the clusters can be found in C1 and C2. We can observe that the clusters of the data from Twitter are less clear than the clusters obtained with the ideas from citizens. Indeed, Twitter data are not focused on citizen ideas about local politics but on all the aspects of life, leading to very noisy data. Nevertheless, by browsing the content of each cluster on the two data sources, we managed to find some relevant ones for the policy-makers, such as:

- Cluster 2 (Democratic life): this cluster summarizes the lack of citizen confidence in politics and solutions suggested to renew the democratic life in Liège (e.g. via the creation an app to follow what happens in the town council, the organization of workshops, etc.).
- Cluster 3 (Local food): this cluster summarizes the willingness of the population to have food that is more locally developed in Liège, through shared vegetable gardens. It also deals with sensitizing the population about eating local.
- Cluster 14 (Social cohesion): this cluster summarizes the ideas and discussions about re-using places in the city to facilitate the discussions between citizens from all ages, origins, and social classes.
- Cluster 30 (Bus and soft mobility): this cluster bundles messages about the bus services, the difficulties of cycling, and citizen ideas on the creation of river shuttles in summer, creation of tracks bicycle routes, adding bus lines, etc.

Of course, less interesting clusters are also identified (related to TV programs for instance) and can be dismissed by the user of the dashboard. The iterative identification of interesting clusters and dismissal of non-relevant ones can be supported by the word cloud (D) of the dashboard that quickly gives a sense of the content of each cluster. The word cloud can also be used, in the **Implementation** stage, to phrase the policies close to the citizens' wording.

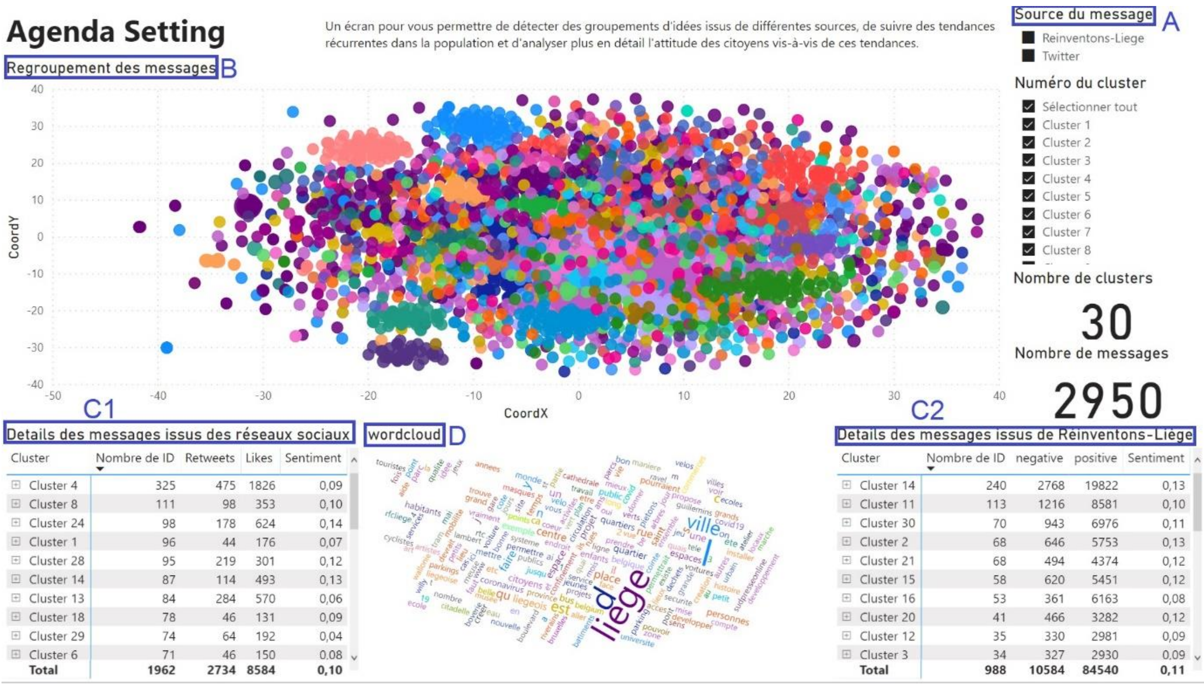

Figure 3. First page of the dashboard

The second page of the dashboard addresses the requirements for the **Formulation** and **Decision** phases, as shown in Figure 4. Regarding Formulation, policy-makers can isolate a previously identified interesting cluster by browsing the list on the left side of Figure 4 and perform another clustering in order to identify interesting sub-subjects or various opinions in the cluster. In this example, we keep the cluster 30 with its content focusing on alternative mobility (mostly bike). The new division obtained divides the text into two sub-clusters: one focused on mobility near the river while the other one tends to focus on natural parks and nature.

It also allows checking the match between the ideas on the platform and the themes discussed on social media. For instance, in Cluster 6, a lot of citizens are giving ideas and complaints about the COVID-19 lockdown but no counterpart was found in the platform. This could allow policy-makers to identify hot topics on social media that are maybe not reflected yet in a category on the e-participation platform. On the other hand, the majority of the clusters draw from messages from both sources. This allows enriching the projects of the platform with comments from social media and reversely translate discussions on social media into concrete projects.

Regarding the **Decision** stage, it provides policy-makers with quantitative insights (E) about the Twitter posts (likes and retweets), ideas from the platform (likes and dislikes) and overall sentiment. It allows policy-makers to rank the posts and ideas based on this quantitative info to facilitate their analysis and present them in a priority order. Furthermore, through the search engine (F), it allows searching for keywords (that could be identified in the word cloud) to facilitate the exploration of the ideas. The targeting of a specific word is ideal to see in which clusters lies a specific word (e.g. COVID, bikes, train, …) appears. The full content of the idea/post is then displayed (G)

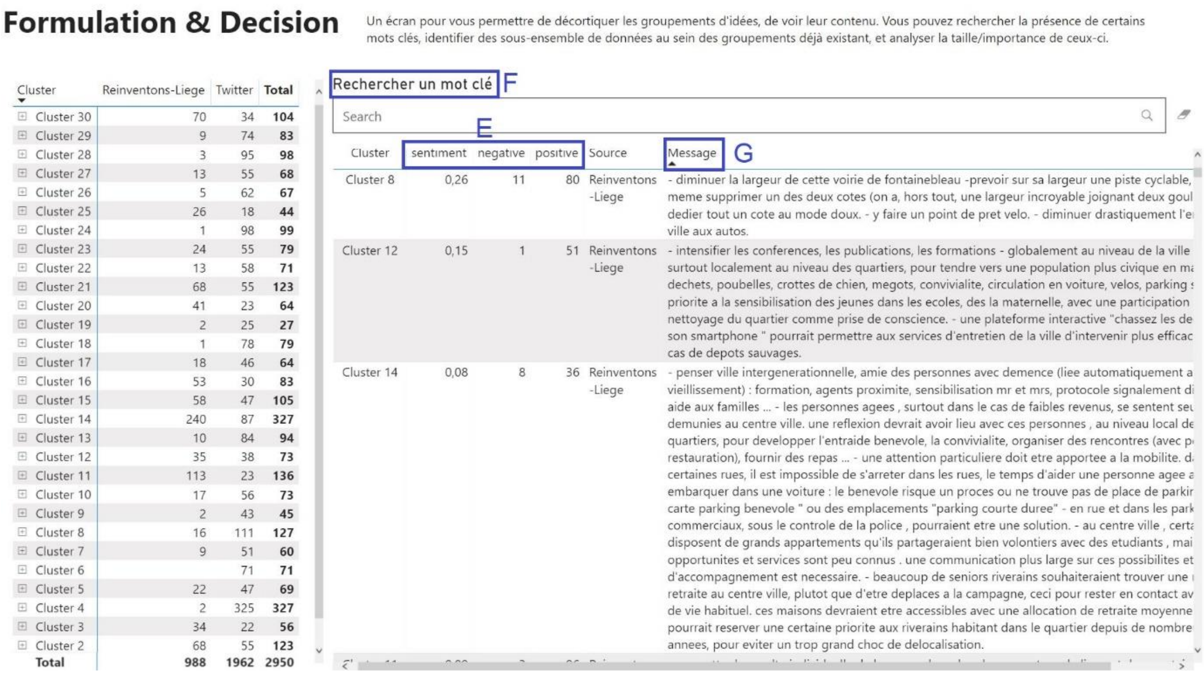

Figure 4. Second page of the dashboard

Lastly, the third page of the dashboard addresses the requirements for the **Monitoring** phase, as shown in Figure 5. Figure 5 shows the evolution of sentiments for clusters 2, 3, 14 and 30. Once a policy has been developed, the policy-maker may wish to investigate its impact on the citizens. They could monitor social media in order to detect a change in the appreciation of the citizens. The resulting visualization allows the policy-maker to monitor the sentiment conveyed by tweets through time.

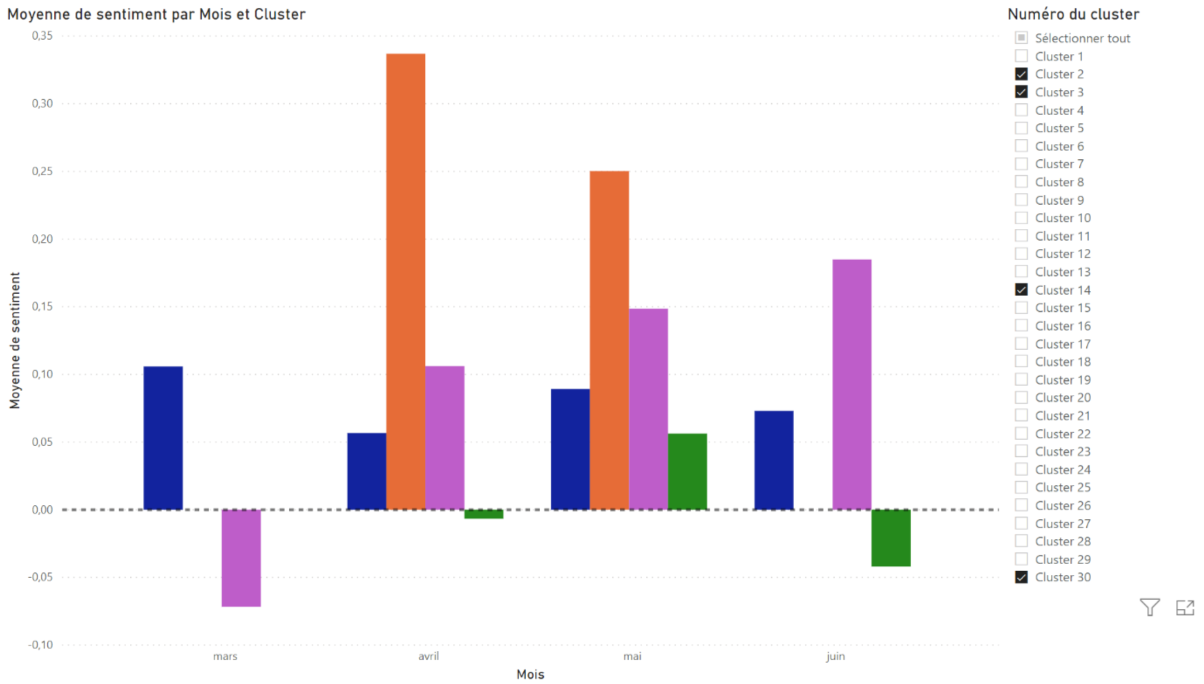

Figure 5. Third page of the dashboard

Finally, in the fourth page of the dashboard, we included a visualization allowing technical users of the dashboard to use the elbow method for identifying the optimal number of clusters to browse the data in total transparency. To apply the elbow method with K-Means, a classical metric called inertia represents the sum of the distances of each document to their assigned K-Means centroid. However, it must be noted the distribution of our data is such that the inertia drops on a straight line making it hard to spot a sharp elbow, despite a decreased drop of inertia between 25 and 35 clusters.

4.4. Dashboard Validation for Liège

In order to validate the dashboard, we performed two follow-up interviews with the two practitioners from Liège. These interviews took the form of a think-aloud evaluation (Hartson, Rex, 2012) where the two practitioners were asked to explore the platform. This method allows users to verbally express their tensions and positive elements about their interaction experience without major intervention from the researchers. The objective is to take the users' perspective. The positive points and leads for improvement collected from the practitioners are presented drawing from the validation themes of Hagen et al. (2019):

- **Relevancy of information**: Both practitioners underlined the relevance of linking unstructured social media posts to the structured information from the platform. However, one interviewee mentioned that the social media posts found in the dashboard were generally less relevant and usable for policy-making.

- Interpretability: The exploratory browsing of the ideas is intuitive and constitutes the main interest of the dashboard. However, they still think a human check of the cluster is necessary. The addition of a search engine was suggested by one interviewee to improve this browsing. The automatic labelling of the clusters (e.g. naming after the most cited term in the word cloud) would have been ideal to interpret the clusters but was not possible due to the technical limitations of Power-BI.

- Learnability: The interface is easy to understand after a brief explanation. As a lead for improvement, one interviewee suggested the addition of small explanations about the purpose of the dashboard on each page to help explore it more efficiently.

- Utility: The two stakeholders mentioned that the dashboard would help them in their daily tasks related to e-participation. The dashboard would especially be helpful at the start of the project to avoid the tedious manual clustering of ideas. The monitoring page of the dashboard constitutes a nice way to see the evolution of citizens' reaction towards a project. If the city enforces the use of a hashtag to follow a project on both channels, this would facilitate the grouping of the clusters and the evolution of citizens' reactions towards this hashtag. Overall, they considered that the analytics features are more holistic than the ones suggested by the private provider of the platform. Two leads for improvement to increase the utility of the platform were not taken into account due to technical limitations. As Belgium is a bilingual country, stakeholders mentioned that the analytics techniques should be applied to Dutch and French. However, we do not have a performance measure about data processing in Dutch. Second, the addition of a button to give feedback to an idea would be ideal but requires access to Twitter or the e-participation platforms.

5. Discussion

5.1. Theoretical Implications

This paper is an early attempt at formalizing the policy analytics process in the context of e-participation. The suggested framework inserts itself at the crossroads of several research fields: public administration (policy-making steps), digital government (policy analytics), e-participation and data science. Our approach draws its robustness from these different approaches.

In the background, we presented the emerging theory-building of policy analytics with seminal studies such as (Gil-Garcia et al., 2018). In this study, we have identified 10 requirements to understand what policy-makers expect from analytics in the context of e-participation. Our framework then contributes to this theory- building by providing a clear mapping of data analysis techniques to each of the five studied steps of policy-making. Furthermore, to the best of our knowledge, it constitutes the first attempt at formalizing a policy analytics process in the context of e-participation.

Despite being one of the most popular participation channel, the use of analytics for platforms remains scarce (most of the studies focus on e-petitioning platforms, Marianne et al. (2019) being a seminal study for the type of platform studied in this paper). Therefore, our investigation of techniques provides another theoretical basis in this context. Moreover, the originality of our framework resides in the mining and combining of data from e-participation platforms with a complementary channel (social media) to support effectively the policy-making process of representatives.

5.2. Practical Implications and Recommendations

The research developed in this paper also has practical implications. First, thanks to the presentation of the data analysis techniques in a dashboard, the framework developed is actually easily usable by practitioners. Furthermore, its inherent modular nature allows for incremental adaptation and improvement over time as the requirements of stakeholders might change.

Second, as mentioned in the Section 3, this framework can be embedded in a broader process as formalized by Cronemberger & Gil-Garcia (2020). Therefore, it constitutes a foundation for a global methodology to be built around the framework. Relying on insights from the interviews and on the policy analytics formalization described by (Cronemberger & Gil-Garcia, 2020), we issue the following recommendations to policy-makers:

- **Governance of the project**: We recommend the project has a clear leadership so that the e-participation process is managed properly to avoid leading to negative outcomes (Lee & Kim, 2014). In the four studied cities, a dedicated "smart city" or "e-participation" manager was identified as the responsible, end-to-end, for the process and constituted a success factor in the projects.

- **Conceptualization of the problem**: We recommend policy-makers to carefully identify the policy issue on which they would like to investigate citizens' opinions. The goals of the participation activities and the impact of citizens on the process should be clearly defined (e.g. following the classification of Arnstein (1969)). We believe the clear conceptualization of the problem will drive the data collection and analytics phases.

- **Data collection and preparation**: We recommend a careful preparation of the data to be inputted into our dashboard (Matheus et al., 2020). The collection of data from Twitter was more challenging then the one from the platform. Indeed, the text from the platform was more structured. However, some data cleaning from Twitter was necessary to remove retweets, special characters, and tweets of less than 200 characters.

- **Data analytics**: The framework presented in this paper inserts itself in this step as we explored different techniques based on the identified requirements.

- **Data visualization**: The dashboard offers visualizations to policy-makers. However, we recommend to investigate this aspect further to explore how visualizations can be used to present the insights from platforms and social media to non-expert citizens directly (in line with Requirement 10).

6. Conclusion

This paper contributes at several levels. First, we identified 10 policy-makers' requirements from four different cities (Liège, Mons, Leuven and Marche-en-Famenne) regarding mining e-participation platforms and social media for policy analytics. These requirements are mapped to the five stages of policy-making. Second, we explored relevant data analysis techniques (document clustering, document filtering, ranking and sentiment analysis) and their combination to address the requirements. Finally, we suggested a policy analytics framework, presented as a dashboard for usability, consisting of a mapping of data analysis techniques to the five stages of the policy-making cycle to address the elicited requirements. This framework supports the whole policy-making process by:

- Identifying recurring tendencies, through document clustering and filtering, from the population to put in the political agenda for the "Agenda-Setting" phase;
- Translating macro-tendencies in concrete alternatives, through document clustering and filtering, for policy in the "Formulation" Phase;
- Presenting a ranking of citizens' ideas and posts from both channels to help the "Decision" phase;
- Using the wording of the citizens when executing the policies in the "Implementation" phase, through a word cloud;
- Examining citizens' reactions to the actions taken by policy-makers, through sentiment analysis, in the "Monitoring" phase.

Furthermore, we provided a preliminary validation of this framework to the use case of Liège. Practitioners underlined the relevance, ease of use, interpretability and utility of the framework to support the policy-making process.

This paper also has inherent limitations and introduces relevant leads for further research. First, out of the 10 requirements identified with the stakeholders, our framework addresses 8 of them. Therefore, we would suggest the interested researchers to investigate further the two remaining and examine alternatives from the 8 studied. For instance, for Requirement 7 (use the wording of the citizens), we would suggest to investigate Natural Language Processing solutions to match the wording of policies to the wording citizens use on both channels. Second, one formalization of the policy-making process around five stages by Howlett et al. (2009) was chosen in this paper. Other formalizations should be explored as well such as (Gerston, 2014) or (Jann & Wegrich, 2017) to check if different data analysis techniques can be matched to more fine-grained stages of policy-making. Third, this paper focused on the use of analytics to support policy-making. However, the use of analytics cannot be considered as a "silver bullet" to automatically improve decision-making, especially to solve complex issues (Ghasemaghaei, 2019). For instance, relying on insights only from social media analytics can highlight polarized point of views and non-constructive argumentation to the decision-maker (Hong & Kim, 2020). Therefore, the combination of analytics and participation methods is a promising lead for further research. In this paper, the processing of the offline participation of citizens was left out of scope.

Future research should focus on where offline participation can support better the decision-making of policy-makers. Furthermore, the mining of information from social media and e-participation platform could in turn fuel offline participation activities. This ecosystem view of offline and online participation to improve policy-making would deserve attention. Fourth, the identified requirements are based on four cities in Belgium. The interview guide used in this paper should be used to collect data from cities with different characteristics in other countries to see if different requirements would emerge. Finally, there is a limitation inherent to our technical expertise and the data we used. Other data analysis techniques are relevant to address the requirements and should be explored in further research. For instance, finding ways to quickly visualize and explore the LDA generated topics in order to leverage the better text classification it proposes or exploring classification based on newly developed language models such as BERT or GPT-3 constitute two promising ways forward. As our framework is modular in nature, the replacement and addition of such techniques is facilitated. Furthermore, other type of participation data should be used for the analysis such as Facebook data.

Acknowledgements

We would like to acknowledge the Belgian Federal Science Policy Office (BELSPO) for their support. The research pertaining to these results received financial aid from the Federal Science Policy according to the agreement of subsidy no. [B2/191/P3/DIGI4FED].

References

Abu-shanab, E. (2013). E-Participation levels and Technologies. In *Proceedings of the 6th International Conference on Information Technology*.

Anderson, R. (2007). Thematic Content Analysis (TCA): Descriptive Presentation of Qualitative Data Using Microsoft Word. *Descriptive Presentation of Qualitative Data*, 1–4.

Androutsopoulou, A., Charalabidis, Y., & Loukis, E. N. (2015). Using Social Media Monitoring for Public Policy Making: An Evaluation. In *Proceedings of the Mediterranean Conference on Information Systems (MCIS)* (pp. 1–15).

Arnstein, S. R. (1969). A Ladder Of Citizen Participation. *Journal of the American Planning Association*, *35*(4), 216–224. https://doi.org/10.1080/01944366908977225

Asher, M., Leston Bandeira, C., & Spaiser, V. (2017). Assessing the effectiveness of e-petitioning through Twitter conversations. *Political Studies Association Annual Meeting 2017*.

Belkahla Driss, O., Mellouli, S., & Trabelsi, Z. (2019). From citizens to government policy-makers: Social media data analysis. *Government Information Quarterly*, *36*(3), 560–570. https://doi.org/10.1016/j.giq.2019.05.002

Bengfort, B., Bilbro, R., & Ojeda, T. (2018). *Applied text analysis with Python : Enabling Language-Aware Data Products with Machine Learning*. *O'Reilly Media*. O'Reilly Media.

Berntzen, L., & Johannessen, M. R. (2016). The Role of Citizen Participation in Municipal Smart City Projects: Lessons Learned from Norway. In *Smarter as the New Urban Agenda* (pp. 299–314). Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-319-17620-8

Bonsón, E., Perea, D., & Bednárová, M. (2019). Twitter as a tool for citizen engagement: An empirical study of the Andalusian municipalities. *Government Information Quarterly*, *36*(3), 480–489. https://doi.org/10.1016/j.giq.2019.03.001

Chowdhury, G. G. (2003). Natural language processing. *Annual Review of Information Science and Technology*, *37*(1), 51–89.

Cihon, P., Yasseri, T., Hale, S., & Margetts, H. (2016). Tweeting for the cause: Network analysis of UK e-petition sharing activity. In *Proceedings of the Internet, Politicy & Politics Conference* (pp. 1–30).

Clay, S. (2011). The Political Power of Social Media: Technology, the Public Sphere, and Political Change. *Foreign Affairs*, *90*(1), 28–41.

Cronemberger, F., & Gil-Garcia, J. (2020). Problem Conceptualization as a Foundation of Data Analytics in Local Governments: Lessons from the City of Syracuse, New York. In *Proceedings of the 53rd Hawaii International Conference on System Sciences* (pp. 52–62). https://doi.org/10.24251/hicss.2020.251

Daniell, K. A., Morton, A., & Ríos Insua, D. (2016). Policy analysis and policy analytics. *Annals of Operations Research*, *236*(1), 1–13. https://doi.org/10.1007/s10479-015-1902-9

De Marchi, G., Lucertini, G., & Tsoukiàs, A. (2016). From evidence-based policy making to policy analytics. *Annals of Operations Research*, *236*(1), 15–38. https://doi.org/10.1007/s10479-014-1578-6

Desdemoustier, J., & Crutzen, N. (2017). *The State of Development of “Smart City” Dynamics in Belgium: A Quantitative Barometer*. Smart City Institute, Liège.

Dolson, J., & Young, R. (2012). Explaining variation in the e-Government features of municipal websites: An analysis of e-Content, e-Participation, and social media features in Canadian municipal websites. *Canadian Journal of Urban Research*, *21*(2), 1–24.

Drever, E. (1995). *Using Semi-Structured Interviews in Small-Scale Research. A Teacher’s Guide.* Edinburgh: Scottish Council for Research in Education. Retrieved from http://eric.ed.gov/?id=ED394990

Fedotova, O., Teixeira, L., & Alvelos, H. (2012). E-participation in Portugal: evaluation of government electronic platforms. *Procedia Technology*, *5*, 152–161. https://doi.org/10.1016/j.protcy.2012.09.017

Gerston, L. N. (2014). *Public policy making: Process and principles*. *Routledge*. Routledge. https://doi.org/10.4324/9781315701387

Ghasemaghaei, M. (2019). Does data analytics use improve firm decision making quality? The role of knowledge sharing and data analytics competency. *Decision Support Systems*, *120*, 14–24. https://doi.org/10.1016/j.dss.2019.03.004

Gil-Garcia, J. R., Pardo, T. A., & Luna-Reyes, L. F. (2018). Policy analytics: Definitions, components, methods, and illustrative examples. In *Policy Analytics, Modelling, and Informatics* (pp. 1–16). Springer. https://doi.org/10.1007/978-3-319-61762-6_1

Gilbert, N., Ahrweiler, P., Barbrook-Johnson, P., Narasimhan, K. P., & Wilkinson, H. (2018). Computational modelling of public policy: Reflections on practice. *JASSS*, *21*(1), 1–14. https://doi.org/10.18564/jasss.3669

Grover, P., Kar, A. K., Dwivedi, Y. K., & Janssen, M. (2019). Polarization and acculturation in US Election 2016 outcomes – Can twitter analytics predict changes in voting preferences. *Technological Forecasting and Social Change*, *145*(August 2019), 438–460. https://doi.org/10.1016/j.techfore.2018.09.009

Guest, G., Bunce, A., & Johnson, L. (2006). How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability. *Field Methods*, *18*(1), 59–82. https://doi.org/10.1177/1525822x05279903

Hagen, L. (2018). Content analysis of e-petitions with topic modeling: How to train and evaluate LDA models? *Information Processing and Management*, *54*(6), 1292–1307. https://doi.org/10.1016/j.ipm.2018.05.006

Hagen, L., Keller, T. E., Yerden, X., & Luna-Reyes, L. F. (2019). Open data visualizations and analytics as tools for policy-making. *Government Information Quarterly*, *36*(4). https://doi.org/10.1016/j.giq.2019.06.004

Hale, S. A., John, P., Margetts, H., & Yasseri, T. (2014). Investigating Political Participation and Social Information: Using Big Data and a Natural Experiment. *APSA 2014 Annual Meeting Paper*.

Hamza, K., & Mellouli, S. (2018). Background on frameworks for policy analytics. In *Public Administration and Information Technology* (Vol. 25, pp. 19–37). Springer. https://doi.org/10.1007/978-3-319-61762-6_2

Hartson, Rex, P. P. (2012). *The UX Book - Process and guidlines for ensuring a quality of user experience*. *Elsevier*.

Hevner, a. R., March, S. T., & Park, J. (2004). Design Science in Information Systems Research. *MIS Quarterly*, *28*(1), 75–105. https://doi.org/10.2307/25148625

Hong, S., & Kim, S. H. (2020). Political polarization on Twitter: Implications for the use of social media in digital governments. *Government Information Quarterly*, *33*(4), 777–782.

Howlett, M., Ramesh, M., & Perl, A. (2009). *Studying public policy: Policy cycles and policy subsystems*. *Oxford University Press*.

Jann, W., & Wegrich, K. (2017). Theories of the policy cycle. *Handbook of Public Policy Analysis: Theory, Politics, and Methods*, *125*, 43–62. https://doi.org/10.4324/9781315093192-11

Janssen, M., & Helbig, N. (2018). Innovating and changing the policy-cycle: Policy-makers be prepared! *Government Information Quarterly*. https://doi.org/10.1016/j.giq.2015.11.009

Kim, G. H., Trimi, S., & Chung, J. H. (2014). Big-data Applications in the Government Sector. *Communications of the ACM*, *57*(3), 78–85. https://doi.org/10.1145/2500873

Kolkman, D. (2020). The usefulness of algorithmic models in policy making. *Government Information Quarterly*, *37*(3). https://doi.org/10.1016/j.giq.2020.101488

Lago, N., Durieux, M., Pouleur, J.-A., Scoubeau, C., Elsen, C., & Schelings, C. (2019). Citizen Participation through Digital Platforms: the Challenging Question of Data Processing for Cities. *SMART 2019 : The Eighth International Conference on Smart Cities, Systems, Devices and Technologies*, (August), 19–25.

Lee, J., & Kim, S. (2014). Active Citizen E-Participation in Local Governance: Do Individual Social Capital and E-Participation Management Matter? In *Proceedings of the 47th Hawaii International Conference on System Sciences* (p. 10).

Liu, T., Liu, S., Chen, Z., & Ma, W. Y. (2003). An Evaluation on Feature Selection for Text Clustering. In *Proceedings of the 20th International Conference on Machine Learning* (pp. 488–495).

Lurie, N. H. (2004). Decision Making in Information-Rich Environments: The Role of Information Structure. *Journal of Consumer Research*, *30*(4), 473–486. https://doi.org/10.1086/380283

Macintosh, A., Coleman, S., & Schneeberger, A. (2009). eParticipation: The Research Gaps. In A. Macintosh & E. Tambouris (Eds.), *Electronic Participation: Proceedings of the 1st IFIP WG 8.5 International Conference, ePart 2009* (Vol. 5694, pp. 1–11).

Macintosh, A., & Whyte, A. (2008). Towards an evaluation framework for eParticipation. *Transforming Government: People, Process and Policy*, *2*(1), 16–30.

Maragoudakis, M., Loukis, E., & Charalabidis, Y. (2011). A Review of Opinion Mining Methods for Analyzing Citizens' Contributions in Public Policy Debate. In E. Tambouris, A. Macintosh, & H.

Bruijn (Eds.), *Electronic Participation: Proceedings of the 3rd IFIP WG 8.5 International Conference, ePart 2011* (Vol. 6847, pp. 298–313).

Matheus, R., Janssen, M., & Maheshwari, D. (2020). Data science empowering the public: Data-driven dashboards for transparent and accountable decision-making in smart cities. *Government Information Quarterly*, *37*(3). https://doi.org/10.1016/j.giq.2018.01.006

Nam, T. (2020). Do the right thing right! Understanding the hopes and hypes of data-based policy. *Government Information Quarterly*, *37*(3), 101491. https://doi.org/10.1016/j.giq.2020.101491

Panopoulou, E., Tambouris, E., & Tarabanis, K. (2010). eParticipation Initiatives in Europe: Learning from Practitioners. In E. Tambouris, A. Macintosh, & O. Glassey (Eds.), *Electronic Participation: Proceedings of the 2nd IFIP WG 8.5 International Conference, ePart 2010* (Vol. 6229, pp. 54–65). https://doi.org/10.1007/978-3-642-15158-3_5

Pantel, P., & Lin, D. (2002). Document clustering with committees. In *SIGIR Forum (ACM Special Interest Group on Information Retrieval)*. https://doi.org/10.1145/564410.564412

Porwol, L., Ojo, A., & Breslin, J. G. (2016a). An ontology for next generation e-Participation initiatives. *Government Information Quarterly*, *33*(3), 583–594. https://doi.org/10.1016/j.giq.2016.01.007

Porwol, L., Ojo, A., & Breslin, J. G. (2016b). Social Software Infrastructure for e-Participation. *Government Information Quarterly*, *35*(4), S88–S98. https://doi.org/10.1016/j.giq.2016.01.002

Rodríguez Bolívar, M. P. (2018). Policy Makers' perceptions about social media platforms for civic engagement in public services. An empirical research in Spain. In *Policy Analytics, Modelling, and Informatics* (pp. 267–288). https://doi.org/10.1007/978-3-319-61762-6_12

Rose, J., & Saebo, O. (2010). Designing Deliberation Systems. *Information Society*, *26*(3), 228–240. https://doi.org/Pii 921698357 Doi 10.1080/01972241003712298

Sandoval-Almazán, R., Luna-Reyes, L. F., Luna-Reyes, D. E., Gil-Garcia, J. R., Puron-Cid, G., & Picazo-Vela, S. (2017). Fostering Citizen Engagement. In *Building Digital Government Strategies: Principles and Practices* (pp. 95–106). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-60348-3_8

Sarikaya, A., Correll, M., Bartram, L., Tory, M., & Fisher, D. (2019). What do we talk about when we talk about dashboards? *IEEE Transactions on Visualization and Computer Graphics*, *25*(1), 682–692. https://doi.org/10.1109/TVCG.2018.2864903

Schossböck, J., Sachs, M., & Leitner, M. (2016). E-Participation Platform Features and Design Principles. In *CeDEM16 Proceedings of the International Conference for E-Democracy and Open Government 2016* (pp. 69–74).

Simonofski, A., Asensio, E. S., De Smedt, J., & Snoeck, M. (2019). Hearing the Voice of Citizens in Smart City Design: The CitiVoice Framework. *Business & Information Systems Engineering*, *61*(6), 665–678. https://doi.org/10.1007/s12599-018-0547-z

Simonofski, A., Snoeck, M., Vanderose, B., Crompvoets, J., & Habra, N. (2017). Reexamining E-participation : Systematic Literature Review on Citizen Participation in E-government Service Delivery. In AIS (Ed.), *Proceedings of the 23rd Americas Conference on Information Systems*.

Simonofski, A., Vallé, T., Serral, E., & Wautelet, Y. (2019). Investigating context factors in citizen participation strategies : A comparative analysis of Swedish and Belgian smart cities. *International Journal of Information Management*, (October 2019). https://doi.org/10.1016/j.ijinfomgt.2019.09.007

Smith, G. F., & March, S. T. (1995). Design and natural science research on information technology. *Decision Support Systems*, *15*(4), 251–266.

Toots, M. (2019). Why E-participation systems fail: The case of Estonia's Osale.ee. *Government*

Information Quarterly, *36*(3), 546–559. https://doi.org/10.1016/j.giq.2019.02.002

Tsoukias, A., Montibeller, G., Lucertini, G., & Belton, V. (2013). Policy analytics: An agenda for research and practice. *EURO Journal on Decision Processes*, *1*(1–2), 115–134. https://doi.org/10.1007/s40070-013-0008-3

Van Der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, *9*(November), 2579–2605.

Zimbra, D., Abbasi, A., Zeng, D., & Chen, H. (2018). The state-of-the-art in twitter sentiment analysis: A review and benchmark evaluation. *ACM Transactions on Management Information Systems*, *9*(2), 1–29. https://doi.org/10.1145/3185045

Appendix: Interview Guide

General Questions

- Can you describe your function within the city?
- How does the policy design typically takes place in your city ? What is your role in it ?
- In general, what kind of information do you need to produce a policy properly? Where do you find it? What is the role of participation ?
- Can you describe what citizen participation means in your city? Which citizen participation projects are implemented in your city ? Do you rely on a plan ?
- Which actors are involved in the participation strategy of the city ?

Information needs from E-Participation Platforms (RQ1)

- Does your city have an e-participation platform ? If so, since when ?
- What is the goal of the platform ?
- Which information would you be looking for on the platform ?
- Do you think this information can be helpful to improve the policies of the city ? How do you use it ?
- Do you think this information has an impact on other participation methods ?
- How do you analyze the ideas/comments/feedback expressed by citizens on the platform ?

Information needs from Social Media (RQ1)

- What is the presence of your city on social media ? (Twitter/Facebook/Instagram, …)
- Which information would you be looking for on social media?
- Are you using a specific hashtag to discuss with citizens on social media ?
- Do you think this information can be helpful to improve the policies of the city ? How do you use it ?
- Do you think this information has an impact on other participation methods ?
- How do you analyze the ideas/comments/feedback expressed by citizens on social media?

Data Analytics and complementarity of the channels (RQ2)

- Do you apply any data analysis techniques to extract information from e-participation platforms?
    - If yes, what are the challenges related to the technique used ? (use of probing questions and examples: data quality, visualization, problem conceptualization,…)
    - If no, why not ? Which technique would you like to use ? How would you like the information to be represented ?
- Do you apply any data analysis techniques to extract information from social media?
    - If yes, what are the challenges related to the technique used ? (use of probing questions and examples: data quality, visualization, problem conceptualization,…)
    - If no, why not ? Which technique would you like to use ? How would you like the information to be represented ?
- Do you think the information found on both channels can be complementary ?
- Do you combine the information from both channels ?
- How would you like to see the information combined ?

Framework development and impact on policy-making (RQ3)

- We are developing a framework to help practitioners deal with the analytics techniques to extract information from these two channels. What would you expect from such framework ?
- Which type of support would you need to exploit the data found on these channels ? Which form would you like it to have ?

Closing questions

- Would you like to address another topic related to the research that we didn't highlight in the interview ?
- Could we access to the data on the e-participation platform from your city for research purposes and test our framework ? We would, of course, be happy to share our framework and the results of our research with you.
- Are there other stakeholders in the city that would be relevant for us to interview (political representatives that use the participation channels) ?
- Would you be willing to perform a quick follow-up interview once the framework is developed ?