inter.noise
27-30 AUGUST HONG KONG 2017

Towards the assessment of community response to noise through social media

Luis GASCO*[1,2]; Cesar ASENSIO*[1]; Guillermo DE ARCAS[1]

[1] Universidad Politécnica de Madrid, Spain

[2] EIT Digital Doctoral School DTC Madrid, Spain

ABSTRACT

Noise is an important concern in modern societies affecting the health and wellbeing of citizens. Beyond the objective physiological effects of noise, noise response has a significant subjective component, which is reflected as a community response and has been traditionally evaluated through surveys. These surveys are often costly, invasive and people do not usually take part in them, whether you use one-to-one interview, phone-based polls or web-based forms.

But the big boost of online social networks has demonstrated that some people are willing to share their views and feelings about everyday problems, including noise. Policy makers should pay attention to these new channels, as they can provide insights about community response and provide new ways of measuring subjective modifying factors in a faster and less expensive way. Online Social Networks act like citizen observatories, and their data can be analyzed as a trustworthy source of information, as humans can contextualize situations and discriminate non-important data.

The analysis of these human-sensor data could give us the raw material to know the community response to noise in cities, and the citizens' views regarding different aspects of noise, or specific sound sources. It can also provide a descriptor of reactions towards the performance of actions against noise, something essential to engage stakeholders and improve the efficiency of policymaking in the future. An automatic process in which noise opinions on the Internet are gathered, clustered and analyzed, being able to provide a subjective evaluation of any noise source can be conceived. Today this is something feasible, and it would suppose a breakthrough approach to noise assessment in cities.

This paper describes possible examples of the potential of this new approach in noise management and the key methodological aspects that should be considered for this aim, such as the processes to follow and the technologies to use.

Keywords: City Science, Online Social Networks, Social Media, Environmental Noise, Community Response

I-INCE Classification of Subjects Number(s): 52, 61, 66, 84, 85

1. INTRODUCTION

In 2006 *Time* magazine named "you" as the person of the year (1). This shows the importance of the human being in the new connected world, where the Internet is used as a platform to share views, feelings and to create virtual networks, providing much information that could be used to improve their quality life. The improvement of the Internet combined with the emergence of new products, such as smartphones, has resulted in new behaviors in the population that began to interact with other people using new online platforms such as Online Social Networks. Furthermore, evolution in technologies has allowed connecting different types of sensors and devices to the Internet, something known as the *Internet of Things* (IoT) or the more recent concept of the *Internet of Everything* (IoE). The combination of these factors has created an environment where millions of data coming from environmental sensors, traffic information or social media are created every day. Today, with the new computation methods and algorithms, this data can be analyzed to get information that could improve the people's quality life.

* luis.gasco@i2a2.upm.es ; casensio@i2a2.upm.es

When we speak about new computation methods, we are not just talking about new cluster-computing frameworks such as Hadoop or Spark, but also about new paradigms in computing such as *Human Computation,* where *human sensors* are included. An example of this approach in acoustics can be found in some projects such as Noisemap and Noisetube (2,3), where the people download a smartphone application and make a sound measurement with their devices; for this reason, this paradigm can also be named as *participatory sensing*.

On the other hand, we have the *social computing* approach, which is an area that joins the social behavior and computational systems and which can provide benefits to the final citizens if used correctly. This approach has already been explored in other research areas such as air pollution, natural disasters monitoring and water management. For example, many studies have demonstrated that social media data can be used to forecast the air-pollution levels (4). Social media can also be used to detect natural disasters quickly and even know where the impact to the population is greatest, something critical to properly manage the available aid (5–7). In acoustics, specifically in soundscapes, the approach of analyzing social networks has been combined with urban science in the Luca Maria Aiello Chatty maps, where analyzing meta tags from Flickr photographs gives information about the primary noise source in Barcelona and London streets (8).

The new policy making models, such as Policy Making 3.0, are conceived as participatory systems where policies are done based on evidence. Franco Accordino says that Policy Making 3.0 is "based on the metaphor of a 'collective brain' "(9). The fact that collective intelligence is one of the areas of Social Computing is not accidental, as it refers to the way of acquiring knowledge, using the data that is offered by many individuals using virtual environments. Today we can implement de mentioned paradigms in noise management systems to create value out of the data generated by noise-sensor networks, city data, and social media. We could now combine the human ability to put into context and comprehends the meaning of different things, such as noise events (something known as situational awareness (10)), with the processing power of machines.

So far, in environmental noise, social surveys have been the main way to get the citizenship participation. An evolution of these social surveys are online social surveys (11). Both approaches have some significant drawbacks, such as the high cost of the development of a survey, the indirect costs in getting a representative sample and the fact that people who participate in the study could be part of an annoyed-group that could bias the final result of the general attitude towards noise. Furthermore, sometimes the number of respondents in an online survey could be less than the impact that the study could have on the Internet and the number of comments about it on social media. That is the type of participation we have defined as "passive participation": it is the participation coming from people who decide not to use the usual channel, but whose data is also valuable to get insights.

The people's participation in social media could also be biased, but in general terms, people tend to speak about what is really important for them. For that reason, we believe that the analysis of social media using the mentioned technologies could be an approach to complement the traditional survey approach but with fewer costs and getting insights about the real feeling about noise problems.

In this paper, we will show some aspects to take into account to do this type of research in environmental noise, as well as some examples of use that could be done using social media data and Smart city data.

2. KEY CONCEPTS IN SOCIAL MEDIA ANALYSIS

The first key aspects when we start doing research on the area of social computing is that most of the data will be images and text data. In this paper, we will focus on textual information. For that reason, the process of analysis is pretty related to text mining operations (12): data retrieval, pre-processing, and analytics.

2.1 Data retrieval

Data is the raw material to make our study. The first step is to find data sources in order to get raw information to analyze. Based on the available Internet sources, we can classify the information as data from social media, data from local Open Data platforms or multimedia content from the Internet, although the latter will not be explained in this paper.

More specifically, in the data retrieval for policy making in environmental acoustics, it would be interesting to get information from the particular social media sources specified in Table 1. The elements of the table are considered social media because the users could exchange their opinions using online platforms as Online Social Networks or use the particular comment sections from blogs,

online newspapers or online forums.

Table 1.Types of social media. Based on (13).

Class	Description
Online Social Networks	Online Social Networks (OSN) are networks where people with same interests can establish a virtual relationship. Facebook, LinkedIn, Twitter, and Instagram are OSN with raw data that could be retrieved using APIs[2].
Blogs	A blog is a website where some individual writes about their opinions and feelings. Data from blogs are usually extracted using web scrappers[3] or APIs.
Online newspapers	Newspapers often publish their news in online web pages where users can interact using the comments sections. Data is retrieved using web scrappers.
Internet forums	An Internet forum is a website where a user can interact with other people using public messages classified by topics. Data from forums can be retrieved, using both APIs and web scrappers.

As mentioned above, we can also download information using the Open Data platforms from smart cities. Here we can find different data related to noise, traffic, pollution or even other useful data that could be used to be correlated with other data sources. Table 2 shows seen Open Data resources for some big cities in the world.

One of the main problems of these platforms is the difficulty of finding common structures in data from different sources. This issue makes the process of pre-processing a complex task, due to the need of having every data source in the same format. The lack of uniformity in the aggregation level of data is also a problem: we find noise levels per day in Madrid, but noise levels per minute in Paris. These differences between sources make it very difficult to make comparisons. Nowadays, there are new initiatives that are seeking for the uniformity of these data structures, by the design of web ontologies for smart cities' data (14).

Table 2. Data provided by different Open Data Platforms worldwide.

City	Name	Traffic	Noise	Air Pollution	Complaints
Madrid	Madrid datos abiertos	✓	✓	✓	✗
New York	NYC OpenData	✓	✗	✗	✓
Barcelona	Open Data BCN	✓	✗	✓	✗
Paris	Paris Data	✓	✓	✓	✗

2.2 Pre-processing and analysis

Data extracted from the Internet is usually unclean. Furthermore, both text documents and social media data are unstructured: this means that after the data is downloaded, we get a plain text that can only be analyzed with previous pre-processing. Before making any analysis, we have to find the way to "structure" it, in order to make the subsequent analysis easier.

Some of the questions we could ask ourselves in a first moment, referring to preprocess data from texts are shown in Table 3. These questions will let us know if we need, or we can get valuable information about the location of the data extracted, the personal information of the users and other aspects, such as aggregation level of the data gathered. Different social media will provide us various types of information. For example, in newspapers we could not get personal data from the user's comments, but we will get them from Instagram or Twitter posts.

[2] Application Programming Interfaces (APIs). APIs are set of methods to systematically access a service given by a web platform. They allow the user to get data without programming in a low level.

[3] Web Scrapping is a technique that allows extracting information from web sites. It usually works with computer, which acts as a human, and enters to different web sites to obtain the html information.

Table 3. Questions to answer before the pre-processing stage.

Parameter	Questions
Location	Do we have the location information from the extracted data?
	Could we obtain the precise information about location analyzing the content?
	Are we able to infer the city/neighborhood? *
	Should we only consider the information with exact locations?
Users	Could we get accurate personal data about the content written by the user?
	Could we get general data from the user?
	Should we consider the effect of the connections in the network of the user in the content he or she shares?
	Are the users enough to be representative in a research study?
General	Should we use all the data available or just a sample?
	Could we find the differences between content related to noise as a personal problem and general information about noise, such as educational events?
	Is the aggregation level appropriate for the study we would like to develop?

*If the precise information is not possible to be recovered.

Once we have defined what we need and what we could get from the raw data, we can start the next pre-processing step. We are going to show a simplified pre-processing stage for a text classification where Natural Language Processing is used. We have to note that the process will vary depending on the analysis we want to do afterward. In this process, we will convert the raw text into something structured, which we can then analyze, looking for the uniformity of the textual content. The diagram of Figure 1, part "2. Text pre-processing" shows this process:

- First, we tokenize the text content. Tokenization is the process of dividing a document into small pieces, knows as tokens.
- Then, we convert every token to lower case for treating each word equally.
- In some cases, we should remove punctuation to decrease the dimensionality of the content, as well as removing some stop words that are not useful due to their little lexical content in the sentence (15).
- We also need to remove URLs or any other type of text that is not useful for our research.
- After that, we make a lemmatization process, which consists on converting the different forms a word, such as plural forms, in its base form to handle it as the same object.
- Finally, we have to tag words with their corresponding lexical category. Using tagged words we can extract features we will use during the analysis process.

We can apply different techniques to get insights of the text. Some of them are:

- **Topic modeling**, a process where users can extract the subject from various text documents.
- **Sentiment analysis**, which consists on determining the attitude of the person who wrote the text towards a particular topic.
- **Text classification,** which is automatically classifying documents into categories based on content.

Depending on the data we got from other sources such as Open Data, we can design experiments that allow us to develop different uses, as shown in next section.

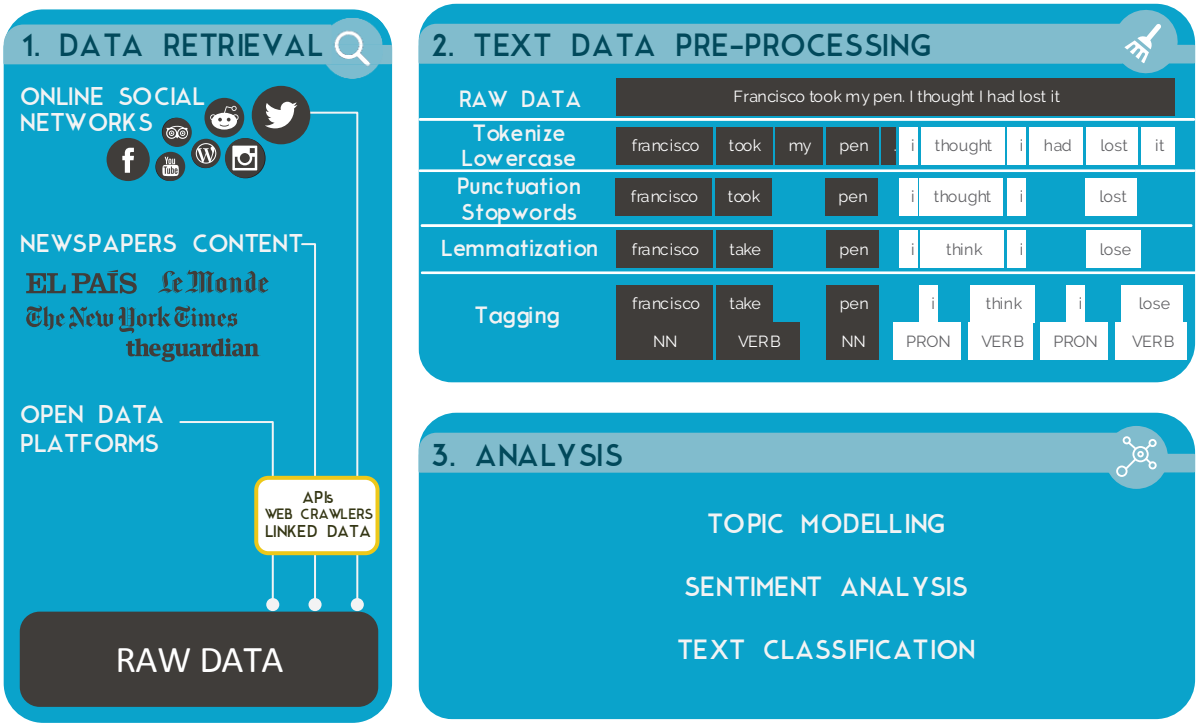

Figure 1. Simplified process to analyze data from the Internet.

3. EXAMPLES OF FEASIBLE USES

The simplified methodology introduces the possibility of analyzing text content from social media. However, which are the possible uses where it is interesting to use these new paradigms? In this section, we present some possible cases of use.

3.1 Noise annoyance

Noise annoyance has a significant subjective component, which is usually evaluated through surveys. With the boost of OSN, people have started to use them to share their views about this problem. If we could analyze these data sources, we could detect noise problems in cities before they become more relevant.

Let us imagine, for example, a scenario A represented in Figure 2, where we could examine OSN to identify people who speak about the annoyance they feel in relation to a particular noise source. We could detect the location of the source that the person refers to or the site of the individual who wrote that content, and we could recognize the degree of annoyance by analyzing the text. Then, we would have a noise source, a set of locations and a group of anonymous people who we know have a noise annoyance feeling. If we plot that data on a map, we could find problems in specific areas of a city, and policy makers/managers could decide if those issues are important enough to be more deeply studied. For example, in Figure 2, we can see two zones with a high number of complaints extracted from Twitter related to leisure noise. Using this methodology, a policy maker could detect these "acoustic/annoyance black points" provoked by this sound source and could eradicate the problem in a shorter period, before it becomes a greater problem.

The idea of monitoring new areas with noise annoyance is useful not only to detect new "acoustic/annoyance black points," but also to see the effects or effectiveness of undertaken actions to mitigate noise in cities, if policy makers used a platform based on these technologies. They could see if the problem has persisted or has moved to other areas. For example, if there are a lot of people drinking in a square and they forbid to drink in there, the number of individuals complaining about it in social media will decrease, but we could detect if the problem has been relocated to other places in the city analyzing OSN.

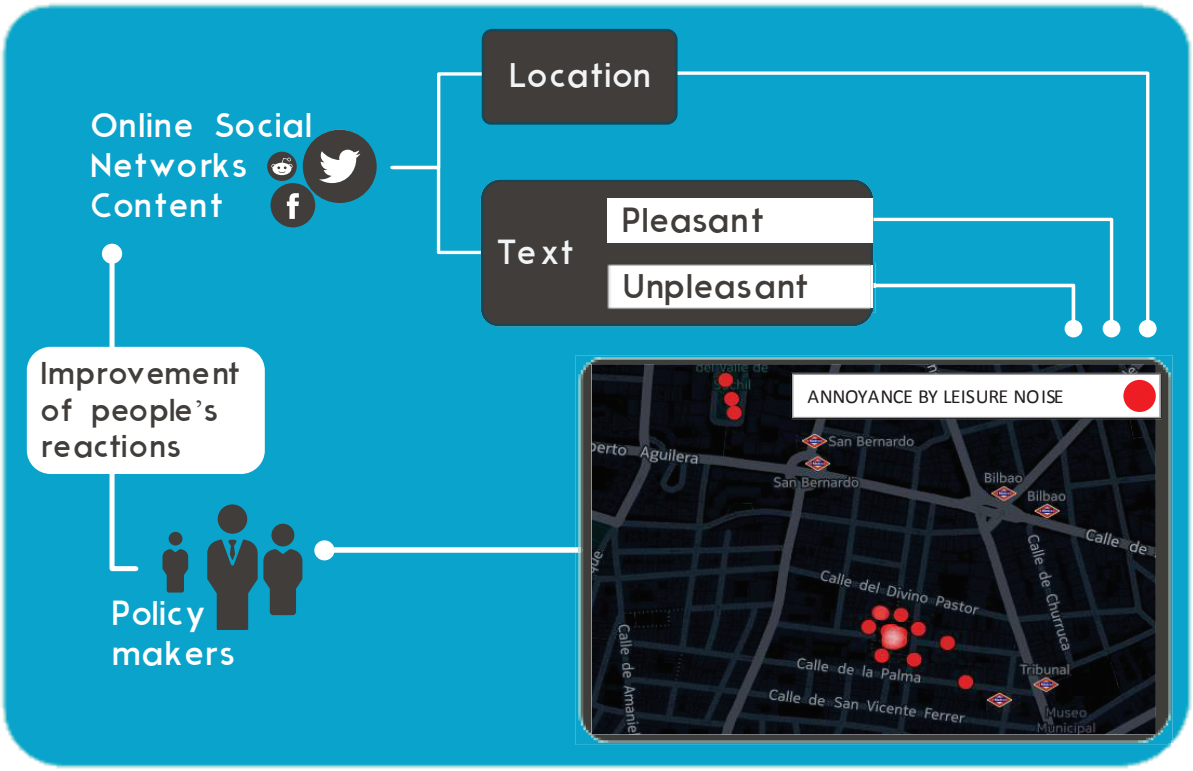

Figure 2. Scenario A. OSN located analysis to improve acoustic comfort in cities.

Noise annoyance is not only related to acoustic factors, but also to many non-acoustic factors, which could affect the community acceptance towards some activities in a city: one of them is the way information is provided to the public(16). Taking that in mind, we can imagine a scenario B where in addition to OSN content, we could get data from smart cities' Open Data platforms. Public data, such as noise monitoring information or noise maps, could be combined with Online Social Networks data, to know if a group of people in a particular area of a city are directly annoyed by high noise levels o if there are non-acoustic factors having an influence in their annoyance.

If noise descriptors were not available, we could infer sound environments (probably low, medium or high noise level) based on road traffic or the number of leisure places (such as pubs or restaurants) in a neighborhood, for example. Alternatively, we could install a noise monitor in those locations to do a particular study.

As an example, in Figure 3, we could explore in depth why in a zone in London, with low noise levels, as seen on the left of the map in the figure, there are a high number of negative opinions on social media such as those of noisier places. With the proposed simplified ideas we present, a researcher or data scientist of a public administration could investigate why that happens. It could be related to the number of pubs in the area, or because there is a bus stop in front of the houses or even due to other non-acoustic factors, as the neighbors do not feel the authorities are paying attention to their noise problem.

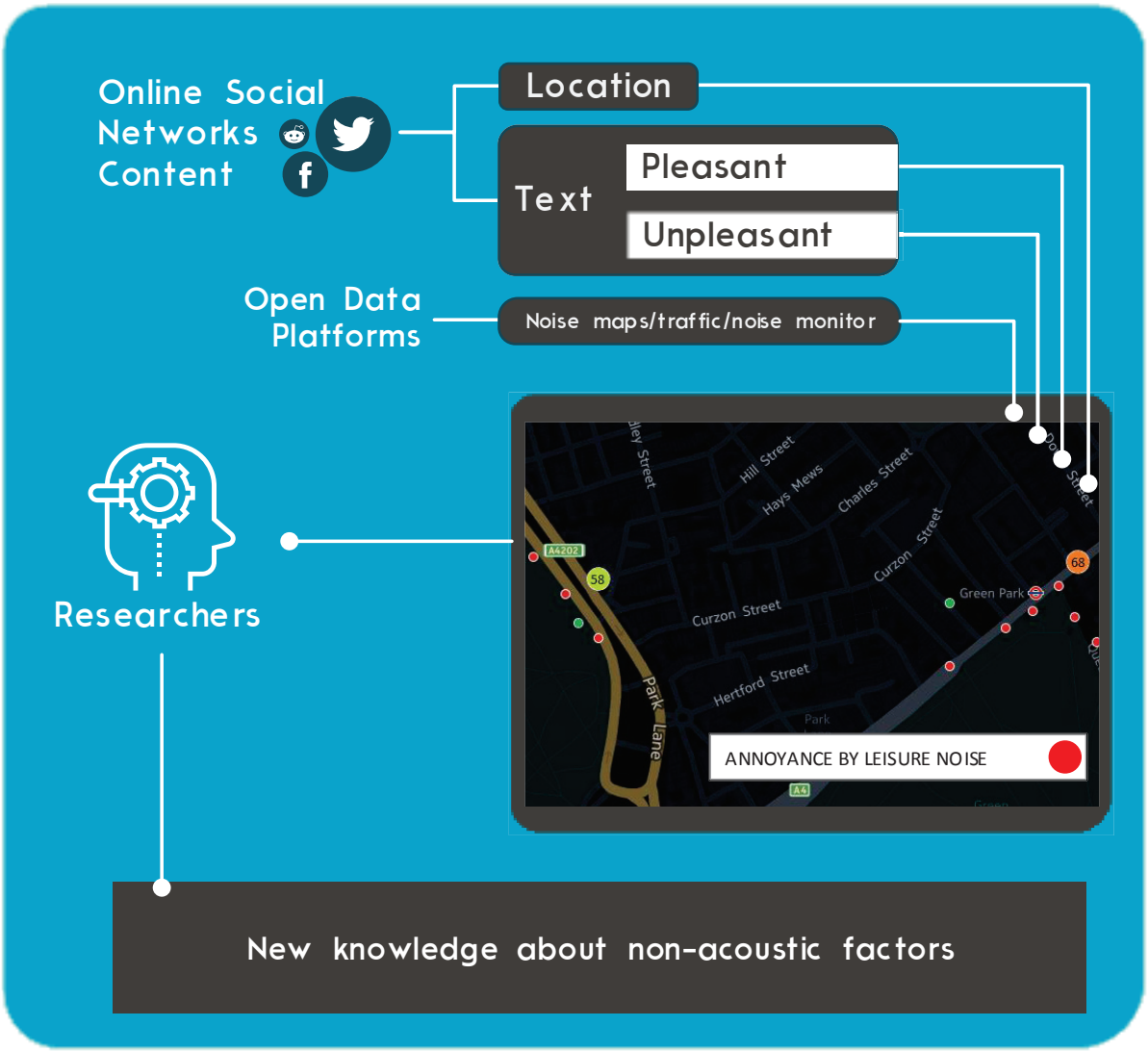

Figure 3. Scenario B. OSN and Open Data Analysis to acquire new knowledge about non-acoustic factors in cities

3.2 Transport infrastructures

Transport infrastructures are one of the main pillars of economic growth and job creation (17). Transport is basic for creating wealth in a city, as it connects different regions, allowing commercial exchanges, people's mobility and job creation to happen. Since 2002, the European Union began to regulate noise pollution, with the aim of controlling and reducing the noise coming from transport, in order to decrease the adverse effects, it has both on people and on the environment. Nevertheless, noise disturbance perception has grown, becoming a problem both for citizens and for the city's policy makers, who feel they cannot keep improving transport systems, for example in terms of public transport or airports, if people are not satisfied with their noise policies.

The above mentioned methodology could be useful for this type of situations. For example, to measure the degree of acceptance of a transport infrastructure, such as a railway station, and how much its noise levels, or even non-acoustic factors such as the way the information is given to the public, affect the citizens' acceptance. Managers could take advantage of this and try alternative methods of disseminating information about their infrastructure, putting the focus in the terms that most affect in the acceptance. If we analyze data sources such as newspaper content and OSN, and we make text mining on it, we could get insights such as "60% speak negatively about the noise coming from railway station in the morning". Then, we could carry out actions such as installing noise barriers in the vicinity of railway stations, in order to increase the acceptance of that specific transport infrastructure.

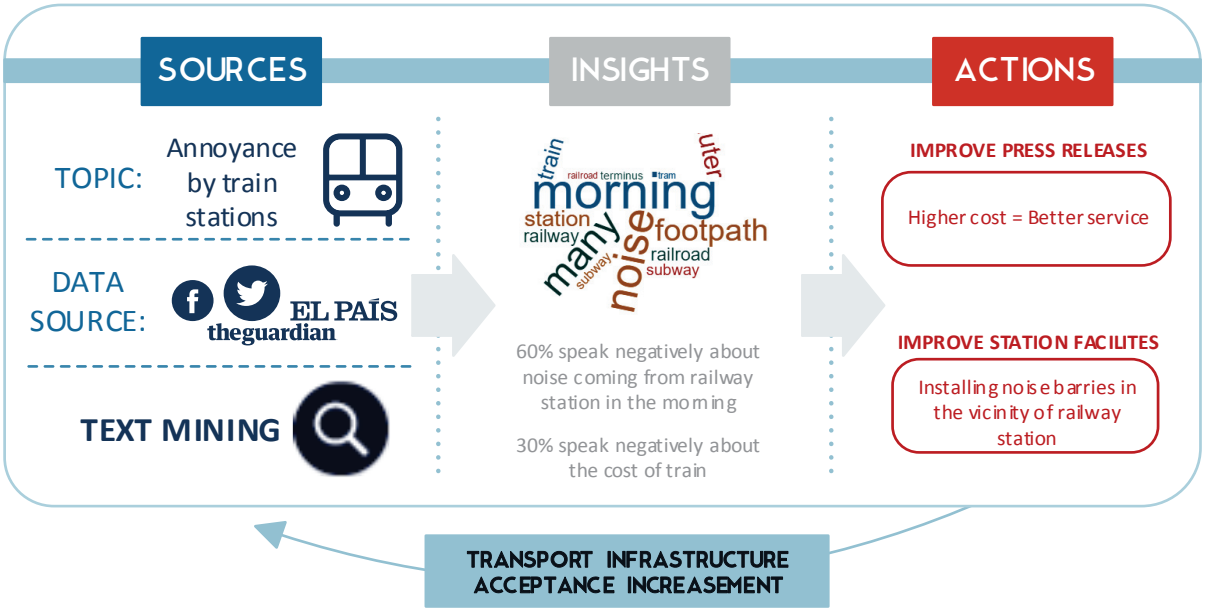

Figure 4. An example of OSN and newspaper analysis to improve transport infrastructures acceptance.

Text mining on social media could also be useful to get citizenship assessments about actions that have not been yet undertaken. For example:

- If the council team expect to close a street or reduce the vehicle speed in a city, these technologies will allow them not only to evaluate the possible noise annoyance changes that will appear, but also other factors such as drivers' reactions or public opinions about the initiatives. The council team will be able to analyze the problems of the new actions, and will be able to tinker them for the common interest.
- If airports managers decide to build a new runway at an airport, news media articles will explain the pros and cons of the construction and people will give their opinion social media. Managers will be able to know the importance of both noise and other factors in the future changes of the airport if they analyze the content of the opinions on the Internet. By knowing the opinions, they could be able to improve the way of making public communications, emphasizing the pros (or what people believe are the pros), and giving better explanations about the cons.

In essence, this is a way of policy-making based on citizens' opinions that work in a cycle. The policy makers or managers of both cities and transport infrastructures could get ideas to improve the city and its infrastructures. After applying it, they should monitor the consequences of those actions in the citizenship, analyzing Social Media Content, and making changes to their initial ideas for the common good.

4. CONCLUSIONS

The Internet, and specifically the social media, is a significant source of data, which we can use to get valuable information for policy makers and city/transport managers. The initial methodology and the examples described in this paper provide a cheaper and auxiliary way to find noise problems in cities and transport infrastructures.

Today, noise risk is monitored using surveys and objective sound level measurements. But citizens are using social media as a channel to express their opinions about different topics, including noise. These views should be analyzed in order to find a relation between the social response and noise effects, as well as between social response and sound pressure levels.

The analysis of these channels, using the technologies introduced in this paper, are part of *policy making 3.0*, as they use the power of the *collective brain* to get insights from public opinions in order to improve the process of making policies adapted to general public.

Using the data from social media and Open Data platforms, we could get more information about

sound environment satisfaction in cities, as well as insights about acceptance of transport infrastructures. We could know why the acceptance of some infrastructures is weak, by knowing the people's opinion in an open and free environment as Online Social Networks are. We could even start new ways of investigating non-acoustic factors, as we could analyze the influence of news media on citizenship and how the way of communicating noise related problems can impact in the annoyance of people.

REFERENCES

1. Sheth AP. Citizen Sensing , Social Signals , and Enriching Human Experience. IEEE Internet Comput. 2009;13(August):87–92.
2. Meurisch C, Planz K, Schäfer D, Schweizer I. Noisemap: Discussing Scalability in Participatory Sensing. Proc First Int Work Sens Big Data Min - SENSEMINE'13 [Internet]. 2013;1–6. Available from: http://dl.acm.org/citation.cfm?id=2536714.2536720
3. Maria E, Maisonneuve N, Stevens M, Niessen ME, Hanappe P, Steels L, et al. Citizen Noise Pollution Monitoring. Proc 10th Int Digit Gov Res Conf. 2009;96–103.
4. Jiang W, Wang Y, Tsou MH, Fu X. Using social media to detect outdoor air pollution and monitor air quality index (AQI): A geo-targeted spatiotemporal analysis framework with sina weibo (Chinese twitter). PLoS One. 2015;10(10):1–18.
5. Al-Saggaf Y, Simmons P. Social media in Saudi Arabia: Exploring its use during two natural disasters. Technol Forecast Soc Change [Internet]. 2014;95:3–15. Available from: http://dx.doi.org/10.1016/j.techfore.2014.08.013
6. Litou I, Boutsis I, Kalogeraki V. Efficient techniques for time-constrained information dissemination using location-based social networks. Inf Syst [Internet]. 2016;64:321–49. Available from: http://dx.doi.org/10.1016/j.is.2015.12.002
7. Le Coz J, Patalano A, Collins D, Guillen NF, Garcia CM, Smart GM, et al. Crowdsourced data for flood hydrology: Feedback from recent citizen science projects in Argentina, France and New Zealand. J Hydrol [Internet]. 2016;541:766–77. Available from: http://dx.doi.org/10.1016/j.jhydrol.2016.07.036
8. Aiello LM, Schifanella R, Quercia D, Aletta F. Chatty maps: constructing sound maps of urban areas from social media data. R Soc Open Sci [Internet]. 2016;3(3):150690. Available from: http://rsos.royalsocietypublishing.org/lookup/doi/10.1098/rsos.150690
9. Accordino F. The futurium - A foresight platform for evidence-based and participatory policymaking. Philos Technol. 2013;26(3):321–32.
10. Chatzigiannakis I, Mylonas G, Vitaletti A. Urban pervasive applications: Challenges, scenarios and case studies. Comput Sci Rev [Internet]. 2011;5(1):103–18. Available from: http://dx.doi.org/10.1016/j.cosrev.2010.09.003
11. Asensio C, De Arcas G, López JM, Pavón I, Gascó L. Awareness: A parallel approach against noise. In: 22nd International Congress on Sound and Vibration, ICSV 2015. 2015.
12. Gandomi A, Haider M. Beyond the hype: Big data concepts, methods, and analytics. Int J Inf Manage [Internet]. 2015;35(2):137–44. Available from: http://dx.doi.org/10.1016/j.ijinfomgt.2014.10.007
13. Daume S, Albert M, von Gadow K. Forest monitoring and social media - Complementary data sources for ecosystem surveillance? For Ecol Manage [Internet]. 2014;316:9–20. Available from: http://dx.doi.org/10.1016/j.foreco.2013.09.004
14. Poveda-Villalón M, García-Castro R, Gómez-Pérez A. Building an ontology catalogue for smart cities. In: eWork and eBusiness in Architecture, Engineering and Construction - Proceedings of the 10th European Conference on Product and Process Modelling, ECPPM 2014. 2015.
15. Martins CA, Monard MC, Matsubara ET. Reducing the Dimensionality of Bag-of-Words Text Representation Used by Learning Algorithms. Proc 3rd IASTED Int Conf Artif Intell Appl. 2003;228–33.
16. Gasco L, Asensio C, de Arcas G. Communicating airport noise emission data to the general public. Sci Total Environ [Internet]. 2017 [cited 2017 May 24];586:836–48. Available from: http://www.sciencedirect.com/science/article/pii/S0048969717303042
17. European Union. White Paper on transport [Internet]. Belgium; 2011. Available from: http://ec.europa.eu/transport/themes/strategies/doc/2011_white_paper/white-paper-illustrated-brochure_en.pdf