Authors final version

AN OPEN DATA CROWDSOURCING APPROACH FOR ENVIRONMENTAL NOISE POLLUTION MAPPING

Iva Cibilić, mag. ing. geod. et geoinf.

Assist. Prof. Dr. Ana Kuveždić Divjak

Assoc. Prof. Dr. Vesna Poslončec-Petrić

University of Zagreb, Faculty of Geodesy, Kačićeva 26, 10000 Zagreb, **Croatia**

ABSTRACT

Open data is undoubtedly one of the most exciting innovations of the last decade. Data that can be (re)used freely and without restrictions is a key driver for achieving many environmental goals. Through the use of open crowdsourcing tools, noise monitoring is enabled by the active participation of citizens who measure noise and thus contribute to the creation of a dynamic noise map. This is consistent with the lifecycle of open data - the design of the process and practices for handling data from its creation, to the provision of open data, to its use by various parties. By participating in the crowdsourcing processes, users act as both consumers and providers of data; with enhanced capabilities of data users, such as commenting, rating, processing, or customizing to their specific needs; and then publishing new versions of the same or their own new datasets. In this way, better communication and collaboration between data users and data providers creates another root for a sustainable ecosystem by closing information and evaluation loops by supporting a broader range of use cases in services and applications on ecological issues.

Keywords: noise pollution, open data ecosystem, environment management, crowdsourcing

INTRODUCTION

Today, noise is defined as any unwanted sound in the environment that disturbs people's sleep, cause them to feel uncomfortable, or have a negative impact on their health. Sensitivity to noise is influenced by the characteristics of the noise (intensity, rhythm, content), individual characteristics of the person (hearing organ condition, age, personal sensitivity to noise), as well as the length, type, and mode of exposure (person's position relative to the source of the noise, presence or absence of noise during breaks taken during working hours and leisure). Outdoor noise is mostly produced by traffic, construction and other public works, industry, leisure, sport, and entertainment. Service units for residential buildings, domestic machinery, and neighbourhood noise are the most common sources of noise in enclosed spaces [1].

A large portion of the population experiences temporary or long-lasting consequences from noise exposure as a result of urbanization and modern living. Considering that the sources of noise are product of human activity, it is possible to act on the overall noise

reduction. Since the noise pollution is an urban phenomenon that occurs in the cities, it is logical that its reduction is the subject of interest of the city governments. In order to address the issue of noise pollution, it is essential to have knowledge about the noise levels around the city as well as the sources of the noise, which is nowadays conducted through noise mapping.

NOISE MAPPING

Noise maps are defined as representations of the current and anticipated level of noise emissions at all sites within the study area, depending on one particular or all sources of noise. Noise control act [2] identifies the measures to avoid, prevent or minimize negative effects on human health caused by environmental noise. It is constructed by creating noise maps based on methods for noise assessing and developing action plan based on the data used in the design of noise maps.

Noise map provides a clear, unmistakable and understandable insight into the problems of noise control. Apart from this, it provides a more effective spatial planning, noise protection planning, and the implementation of "acoustic area zoning" in compliance with legally acceptable noise levels. Environmental noise management system is set up with strategic noise maps, action plans and conflict noise maps.

Strategic noise maps are ones that depict an individual source of noise, i.e., road, rail and air traffic, industrial activity, including marine and river traffic and the recreational facilities (Figure 1).

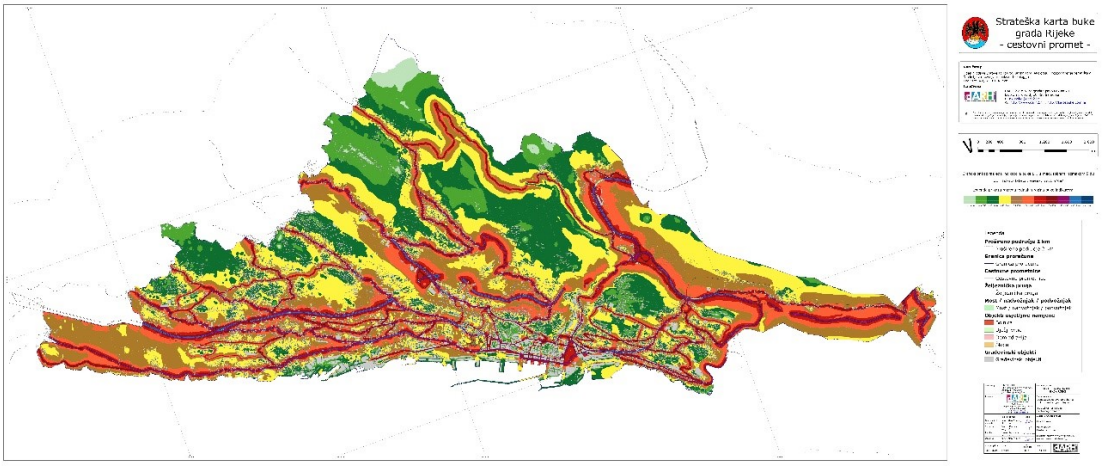

Figure 1 Strategic noise map of road traffic of the city of Rijeka [3]

Act relating to the noise mapping in Croatia is the Regulation on the method of creating and content of noise maps and action plans and on the method of calculating allowed noise indicators (NN 75/09) [4]. This act is in the accordance with the European Union guidelines on the computation methods for calculating noise in industrial areas, major roads, major railways, and major airports. The Act also governs the creation and content of action plans, conflict noise maps, and strategic noise maps.

A conflict noise maps are difference noise maps that are meant to produce action plans [5]. Action plans are plans related to managing environmental noise and its negative impacts, including noise protection measures.

A noise map must generally meet certain requirements in order to be used as a basis for systematic and uniform knowledge acquisition about noise pollution in urban areas. First and foremost, the quantity of data gathered must be sufficient, redundant, and evenly distributed throughout the measuring region. Additionally, data must be obtained throughout a specific period of time. This requires the city government to use new methods of communication and interaction with citizens to activate citizen engagement.

CROWDSOURCING

With respect to ecological issues, many institutions lack the resources necessary to manually sift through large collections of unstructured data. By engaging external communities to collaborate on large collections of information, it is possible to create more detailed machine-readable data that supports a broader range of reuse cases. This process is referred to as crowdsourcing. It is defined as a process in which the work of collecting a large amount of data is broken down into small tasks performed by volunteers. It offers a remarkable opportunity by providing a high scale, on-demand, and low-cost pool of geographically dispersed labor to complete complex tasks.

By definition, crowdsourcing is the method of outsourcing work that would typically be conducted by a chosen individual (often an employee) to a more generalized, wider group of people through an open call [6]. The term was created by the journalists Jeff Howe and Mark Robinson of Wired magazine in an article on the subject in 2006, and as a result, it is now widely used.

The introduction and widespread use of mobile phones have elevated crowdsourcing to unprecedented heights. Smartphones are fairly reasonably priced and come with a variety of built-in sensors, including a camera, GPS receiver, microphone, proximity sensor, compass, accelerometer, gyroscope, and others. With the use of sensors, a range of information may be rapidly and cheaply obtained, and as technology advances, information will become more exact and varied. The possibilities of acquiring a significant amount of data that calls for the usage of sensors is the key driver behind involving people with their devices. The price of the data is another factor that influences the development of open datasets.

OPEN DATA ECOSYSTEM

Open data is currently a hot topic. Open data can be defined as data that is freely accessible online, available for reuse without technical restrictions, and made available under an open access license that allows unrestricted reuse of the data, including across different "scopes of work" (e.g., both commercial and non-commercial) [7].

Various terms have been proposed to describe different models of open data, such as the open data lifecycle, the open data value chain, or the open data process [8]. The open data lifecycle is a conceptualization of the process and practices of dealing with data, starting from its creation, through the provision of open data, to its use by different parties. There are two distinct categories of users in the open data lifecycle:

- Data user (or consumer) - direct data user who performs data processing, analysis and visualization. This user can be a business person, a citizen, or a public official.
- Data provider (or publisher) - performs curation; engages in storage and delivery of data in appropriate formats and metadata standards. This user may be a public entity or a company that publishes data.

To fully realize the benefits of open data, this traditional "one-way" practice of open data can be replaced by an open data ecosystem, i.e., an approach to open data that focuses not only on the accessibility of the data, but also on the larger environment for using open data - its "ecosystem" [9]. An open data ecosystem can be defined as a cyclical, sustainable, demand-driven, and environment-oriented system of actors that are interdependent in creating and delivering value from open data [10]. The added value of the ecosystem perspective on open data is that it focuses on the relationships and interdependencies between the social (publishers and users of open data) and technological (data linkage, Big Data analytics, storage, visualization) factors that influence the performance of open data activities within the lifecycle [11]. For a complete list of methods and tools in each step of the open data ecosystem, see Table 1.

Within an ecosystem, two different sides of the open data lifecycle have been merged into a more comprehensive model to create a better relationship between the above participants. This new type of users is characterized by the duality of their roles, acting simultaneously as both consumers and providers of data (Figure 2). These users may be individuals who have advanced data user capabilities such as annotating, rating, processing for enhancement, customizing to their specific needs, or connecting to other datasets (public or private); and then uploading/publishing new versions of the same or even their own new datasets.

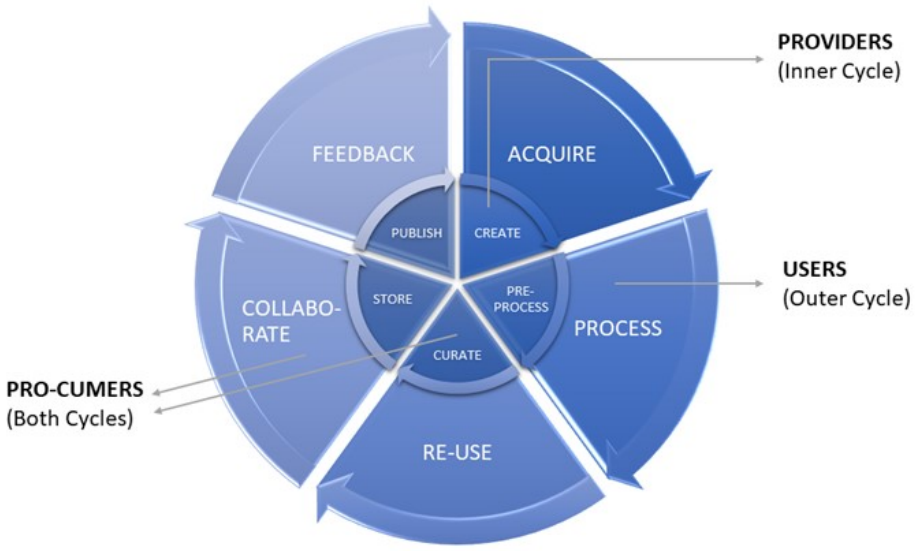

Figure 2 The open data ecosystem, modified after [12]

In this way, another root of a sustainable ecosystem is created through better communication and collaboration between data users and data providers, closing information and valuation feedback loops. The term 'procumers' of data (as a combination of the English words producer and consumer) is introduced, who are both providers (publishers) and users (consumers) of open data, thus improving the number and quality of available open data sets.

Table 1. Methods and tools in each step of the open data ecosystem [6]

Life cycle stage	Tools	Methods
Create/Gather: The process of creating data	Sensors; RFID, IoT, IS; Human; Connection with already gathered open data; Hadoop for big data	Automated data creation (logs, network data); Manual data entry; Linking with Open Data Portals
Pre-process: The managerial process of defining data quality	Detailed Metadata Standards; Evaluation Metrics and Models; Maturity Matrices; Unique identification (URIs and URLs)	Conceptualization & Goal setting; Evaluation plan and data quality; 3-layer Metadata Schema for portals
Curate: The process of meeting the required data quality and legal requirements	LOD Refine External Tool; Individual/Native Tools; R	Structuring; Anonymization; Metadata Refinement; Change Data Format; Data Cleansing
Store/Obtain: The decision making process of storing.	Data Centres; SPARQL Repositories for linked data; NoSQL & Document Databases for big data, linking with other datasets	Versioning; Data Linking; K-value and column oriented databases for big data
Publish: The process covering legal issues	Upload Capability	Publication Plan Open Access Licensing Intellectual Property Rights
Retrieve/Acquire: The process of data acquisition through OD portals	OD portals (e.g. European data portal, world bank, national initiatives)	Multilingual search techniques APIs
Process: The process of data analysis	External data processing tools: Open Refine; R; Rapidminer; KNMINE; excel; Weka/ Pentaho	Data enrichment; Create Linked Open Data; Different Datasets combination; Text and Data Mining; Hashing; Cluster Analysis & Factor Analysis
Use: The process of presenting the analysis outcomes	Internal & External Visualization tools; Statistical Packages; Linking with external artefacts (publications)	Statistical Analysis; Map Visualization; Chart Visualization; Plot Visualization; Visual Analytics; Cluster diagrams
Collaborate: The process of communicating with other data users	Collaboration space and workflow Web 2.0 capabilities and tools	Exchange notes/emails/ideas Create Groups of common interests

Feedback: The process of evaluating and providing feedback to data providers	Declare Need Web 2.0 Capabilities and Tools	Data Quality Rating; Requests on Open Data; Assessment of Publication

ECOSYSTEM APPLICATION IN NOISE MAPPING

The open data life cycle is a conceptualization of the process and practices around handling data, starting from its creation, through the provision of open data to its use by various parties. This process can be explained through the creation of dynamic noise maps by the students of the geodesy and geoinformatics at the Faculty of Geodesy, University of Zagreb.

The process of creating a voluntary noise map includes data collection (spatial, temporal component and noise intensity), data processing (sorting and analysis of collected data) and visualization (classification and production of a thematic map) (Figure 3). Data collection was performed with the mobile application. Since the noise levels in decibels dB (A) and coordinates (WGS84) are available for each data, it proved to be very good and reliable for this task. Each smartphone model has to be calibrated to de-crease the microphone's error in order to produce a relatively accurate data set. The accuracy is estimated for around 5dB. Data processing and final visualization were performed in GIS programs.

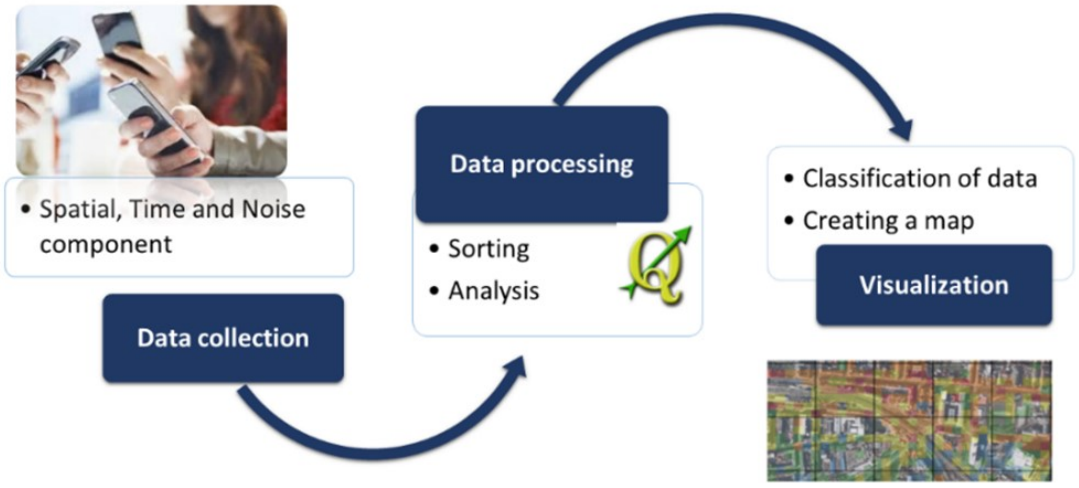

Figure 3 The process of making a volunteer noise map [1]

This project relies on motivating users by offering a free smartphones application to usage (which can also be used for personal purposes without sharing information with the community) and every user that is a citizen in a larger city encounters with noise on a daily basis. This way the awareness of this problem increases among the affected citizens, and together they can participate in its resolution. Relevant authorities and specialists should have comprehensive information on the spatial and temporal dispersion of noise in order to create a strategic noise map. Through the appliance of the ecosystem approach, the producers of the data can also be the ones who analyse it and use for further actions. In this manner, the prodcumers are the ones who can make a significant difference in the dispersion of noise pollution problems among the community and local government.

CONCLUSION

Open data ecosystems perform data production and usage-cycles with feedback loops, sharing of data back to publishers and also with the consumers. Such mechanisms can help users to obtain insight in how they can use and interpret open government data and generate value from them. Crowdsourcing can be an efficient way to increase the quality and availability of open data through the ecosystem cycle model. At the policy level, identifying community crowd-sourced projects outside of government institutions can also be an indicator of valuable datasets that should be prioritized for open release. Noise is one of the biggest pollutants in modern cities, but the risk is often overlooked despite being associated with ranging impacts on the activity of human life. Producing official noise maps and maintaining them is a very demanding and expensive task that falls under the purview of local government. The noise map provides insight into noise management problems and gives a clear, unambiguous, and easy-to-read picture of these problems, which allows for more efficient noise control planning of existing areas and implementation of "acoustic zoning" in accordance with legally permitted noise levels. The model of crowdsourced dynamic maps described in this paper is also applicable to other pollution such as air or light pollution, that can cause adverse effects and degrade environmental quality.

ACKNOWLEDGEMENTS

This research is part of the Twinning Open Data Operational project that has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 857592.

REFERENCES

[1] Poslončec-Petrić, V., Cibilić, I., and Frangeš, S. Crowdsourcing Application in the Development of a Dynamic Noise Map. Proceedings of the International Symposium on Innovative and Interdisciplinary Applications of Advanced Technologies, pp 676–683, 2021. https://doi.org/10.1007/978-3-030-90055-7_54.

[2] NN 14/21: Law on protection against noise. National gazette, No. 14, 2021.

[3] Strategic noise map of road traffic of the city of Rijeka. Available online: https://www.rijeka.hr/wp-content/uploads/2016/07/Strate%C5%A1ka-karta-buke-cestovnog-prometa-za-dan.pdf (accessed on 18 November 2022).

[4] NN 75/09: Regulations on the method of creating and content of noise maps and action plans and on the method of calculating allowed noise indicators. National gazette, No. 75, 2009.

[5] NN 05/07: Rulebook on the method of preparation and content of noise maps and action plans. National gazette, No. 5, 2007.

[6] Howe, J. Crowdsourcing: Why the power of the crowd is driving the future of business, 2008, USA, Random House Business.

[7] Open Knowledge Foundation. Open DataHandbook. Available online: http://opendatahandbook.org/guide/en/what-is-open-data/ (accessed on 20 November 2022).

[8] Welle Donker, F. and Van Loenen, B. How to assess the success of the open data ecosystem?. International Journal of Digital Earth, 10(3), pp 284–306, 2017, https://doi.org/10.1080/17538947.2016.1224938.

[9] Pollock, R. Building the (Open) Data Ecosystem. Available online: https://blog.okfn.org/2011/03/31/building-the-open-data-ecosystem/ (accessed on 20 November 2022).

[10] Heimstädt, M., Saunderson, F., and Heath, T. From Toddler to Teen: Growth of an Open Data Ecosystem. JeDEM - eJournal of eDemocracy and Open Government, 6(2), pp 123–135, 2014. Available online: https://jedem.org/index.php/jedem/article/view/330/284 (accessed on 22 November 2022).

[11] Dawes, S. S., Vidiasova, L., & Parkhimovich, O. Planning and designing open government data programs: An ecosystem approach. Government Information Quarterly, 33(1), pp 15–27, 2016, https://doi.org/10.1016/j.giq.2016.01.003.

[12] Charalabidis, Y., Zuiderwijk, A., Alexopoulos, C., Janssen, M., Lampoltshammer, T., Ferro, E. The Multiple Life Cycles of Open Data Creation and Use. The World of Open Data. Public Administration and Information Technology, vol 28. Springer, Cham., 2018, https://doi.org/10.1007/978-3-319-90850-2_2.