ENVIRONMENTAL DATA SCIENCE SESSION

THE CHALLENGES OF QUALITATIVE DATA SYNTHESIS FOR SOCIO-ENVIRONMENTAL ASSESSMENTS

Filip Alimpić*,
Mesud Adžemović,
Jelena Milovanović

Environment and Sustainable Development
Singidunum University,
Belgrade, Serbia

Abstract:
Available data regarding environmental sciences represents an extremely complex set of inputs since that field of science became one of the most dynamic areas of scientific research. Data science has a huge impact on environmental analysis, offering an abundant scope of new techniques that are used to understand even the most seldom environmental processes. There is no doubt that it is a multidisciplinary challenge because it deals with many fields such as statistics, artificial intelligence, social sciences, psychology, economics, health, etc. Environmental studies forced societies to try to answer questions regarding processes of climate change, human pressures, negative impacts on land, water, and air quality, energy efficiency, health, and food safety, biodiversity and many other aspects. In order to answer those questions, the results of qualitative data analysis must be taken into account when assessing the situation and defining measures for the improvement of particular ecosystems. The aim of this paper is to emphasize a necessity for qualitative data collecting and analyzing through presentation of some assessment examples aimed to define importance of social dimensions in environmental sustainability.

Keywords:
multidisciplinarity, processes, social dimensions, environmental sustainabilityt.

1. INTRODUCTION

Data science is not mainly focused on understanding and managing analyzed data in the environment. Many processes and challenges in environmental science need new techniques and ways to process data. Especially when the most critical topic in today's world is climate change and its impact on life in general. This paper has a goal to determine the potentials of using data science in environmental research, and additionally to highlight crucial challenges for qualitative data analysis in environmental science.

A metonym for data science is a “big data”, which gives us a complex data connotation. However, the real test is when dealing with a diversified/heterogeneous source of data. There is a great potential for data science in the environmental field, and that potential is filled with challenges that can be overcome with new techniques and approaches [1].

Correspondence:
Filip Alimpić

e-mail:
alimpic.filip@outlook.com

DOI: 10.15308/Sinteza-2020-65-69

Understanding the environment represents a multidisciplinary task which includes ecologists, hydrologists, pedologists, biologists, chemists, physicists, and analysists, along with political science which enhances the impact of social sciences in the processes. How big of an impact social science have, can be shown by an example of natural capital and ecosystem services [2].

Natural capital is mainly focused on water, soil, air, energy resources, and living beings, while ecosystem services are primarily focused on sustainable and integrated management of complex ecosystems and all of their aspects, including ecological, social and economical [3].

This shows how much environmental science depends on data [4] (like most other science fields), which leads to a transparent and exact science approach.

It is essential that environmental science effectively integrates data science in its work process, and to look up to branches of science which are more advanced in collaborating with data science, e.g. physics [5] and bioinformatics [6].

These suggestions open up a lot of questions and can start a lot of debates. Most of those questions are related to the integration of the complex sets of data, which have to be transformed in an agreeable collection of new findings which can be later applicable in practice. Most of the practice includes strategies, frameworks, policies, management plans and possible solutions for a series of challenges that environmental science deals with every day.

It is obvious that the relationship between data science and environmental science should be strengthened, and there is no doubt that they should become a symbiotic entity.

2. HETEROGENEOUS DATA CHALLENGES

Challenges and obstacles in data usage are different and can depend on the area of the science studied, however, the main factor which determents that level is diversity in main sources, along with the type of the data involved. Data science most often is identified with data's four Vs: volume, velocity, variety, and veracity [7].

Data from environmental science comes from many different sources, and the processes are supported by many innovations during the collecting:

1. A large number of data is collected through a remote sensor, by observing the environmental area without interfering in one's life flow. It is usually done by using satellite imaging, airplane imaging, and pilotless aircrafts imaging. These types of research include passive activities such as photo and infra-red images, and active actions like LIDAR or Radar, LandSat NASA, and many others [8].
2. Systems and technologies for observing the Earth which uses sensors for collecting data from all layers of the atmosphere: lithosphere, biosphere, hydrosphere, and cryosphere. System based on these technologies is occasionally obtaining the data, although today more often is a real-time obtaining through long-distance transfer. Internet of Things (IoT) is playing a big role in observing the natural environment by using the sensors for real-time data [9].
3. Data obtained through fieldwork is kept growing. The precision of the collected data during the field trip is high in accuracy and a good example is a mobile application in a field of biodiversity called Big Garden Bird Watch [10].
4. Historical records are of great importance for environmental sciences because many of them are digitalized. However, there is still a lot of data that is not digitalized especially on the local level.
5. The analytical approach in the research regarding modern ecology is extended through web platforms such as social media. Many examples show how social media can be used to inform the community about potentially dangerous events like floods, severe weather storms, and many other natural disasters.

Reliability of the data in environmental science is under a question mark. That is the case, especially when dealing with the data collected from local communities. That kind of data can differ based on the common knowledge of the participants of the survey. Also, it is very important to mention that sometimes data obtained through satellite imaging can be less accurate than the data obtained during the fieldwork. Another problem arises from cheap equipment which becomes less and less reliable but easy to acquire.

Diversity of the data in environmental science indicates that we need a way to creatively use data sources, on the other hand, it urges us to connect different sets of findings. The challenge of data in environmental science is the most prominent when one's trying to reach a higher level of the interfering of a couple of sources during the data collection.

Eionet - European information and observation network, which works alongside with EEA (European environmental agency), has a great role in Serbia for integrating the monitoring and reporting on the field. The EEA web of the agencies is made of many European centers, 900 experts from 37 worldwide countries, over 300 national agencies along with other entities which deal with environmental issues.

One of the most significant focuses of information systems is collecting data based on indicators in the environment and their impact on human health, as well as connecting with the World Health Organization (WHO) and their Environmental Health Information System (ENHIS). One of the goals of that department of WHO is to estimate the effect of climate change on human health as well as creating protection measures.

ENHIS is mostly focused on the connection between human health and environmental issues. One of the greatest challenges in this area of interest is the fact that indicators of the status of the environment can differ from the general health conditions of the population.

There are potentially useful, but insufficiently used, qualitative data collected through surveys conducted within local communities with the aim of assessment of social issues relevant for environmental protection and human health [11].

3. QUALITATIVE DATA ANALYSIS EXAMPLE

Research conducted was focusing on professional opinions regarding ecological and economical instruments along with systems of environmental protection. The research was carried through with online surveys sent to 300 e-mail addresses. In the end, 236 surveys were returned filled [12].

Participants are representatives of local government offices, civil society organizations, scientific and research institutions in Serbia.

All the data was collected using Google Forms and stored into Google tables. Cookies were used to ensure that one IP address can deliver one survey.

Questions with numeric (ranking) answers were analyzed in Microsoft Excel, while all answers on open-ended questions were analyzed in NVivo 12 Plus[1] software. It was a useful tool for presenting all kinds of answers for open-ended questions. In that software, answers were coded into categories and analyzed based on the obtained results.

1 NVivo 12 Plus - https://www.qsrinternational.com/nvivo/nvivo-products/nvivo-12-plus

For example, one open-ended question in this survey was: *Who should finance environmental protection*?

All the answers were coded in a category relevant to respondents' opinions. In the end, there were seven categories including government, polluters, local community, non-governmental organizations, international organizations, all together and others.

Fig. 1. shows the results for this question, analyzed using NVivo 12 Plus software and presented as a word cloud.

The analysis shown above cover only one open-ended question. Not only that it presents us with the complexity of the research, but also it illustrates how difficult it is to plan appropriate measures for improvement in the field of environmental protection.

Fig. 1. -Who should finance environmental protection?

The analysis of quality data show that 71.6% of stakeholders think that the expenses for the protection of the environment should be financed by all members of the community (polluters, government, local communities, international organizations, NGOs and others).

Other questions from the survey, which are not shown in this publication, clearly point out the fact that this answer is far from reality. Financial support is currently obtained from polluters and the government for most of the environmental issues. In order to declare responsibilities for each individual to finance protection of the environment in exchange for gaining profit through certain production or processing, it's necessary for quantitative data of environmental impact to exist, especially in the area where such action is happening. Monitoring of the environment is a necessary part of the process of determining the financial obligations of certain parties for the needs of protection of the environment.

On the other hand, existing data regarding the status and intensity of pollution, which are emitted by certain parties, (even the ones who are pretty well systematized) does not necessarily mean much, especially if the local community did not get them serious and accepted them as relevant. The current government stand is that parties who pollute the most should be the ones who finance the protection of the environment.

However, local communities, civil sector and scientists all agree that financing for the protection of the environment should be the responsibility of everyone in that community or country. The ultimate goal of each research is to create a summary for policymakers, which will give a proper pathway for them, but only if they are based on integrations of quantitative and qualitative data of socio-environmental researches.

4. QUANTITATIVE DATA ANALYSIS EXAMPLE

For the purpose of getting better insights on how different reality can be, quantitative analysis was conducted with questions which include ranking series. These kinds of questions were able to draw the answers which will eventually show the real difference between subjective feelings and objectives of the reality.

Quantitative data analysis shown in the *Fig. 2.* included participants from many different educational backgrounds. From individuals with secondary education to individuals with Ph.D. degree.

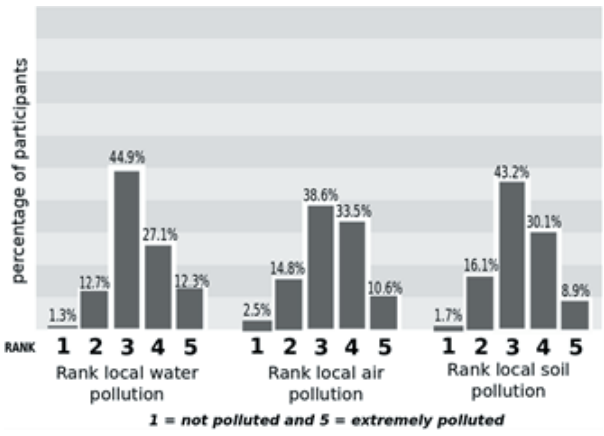

Fig. 2. – Participants ranking the water, air and soil pollution, in their own opinion.

Results were, as expected, an overlap between subjective feelings and objectives of the reality. As it can be inferred from the *Fig. 2.*, greater percentage of participants ranked water, air and soil pollution to be in the higher level of 3 and 4 on a scale of 5. On the other hand, just a few participants marked levels of 1 on the scale, and levels 2 and 5 were marked approximately in the same percentage.

These results were really important to show the awareness of the people in Republic of Serbia, regarding environmental protection. It shows just how much global influence through mass and social media sculptures the minds of society. However, information given by the public institutions is most of the time rare and with a lack of details. Field of environmental studies should focus more on improving the channels for distributing valuable and accurate information regarding, if possible, real time data for all people interested in current status on environmental issues, whether those individuals are environmental professionals or citizens.

There is a huge data, measured from different environmental branches, but most of those results are not analyzed and therefore the real picture and status of the environment can not be established.

Fortunately, the difference between today's assumptions regarding status of the environment and real-time results is not significant. Still, there needs to be better cooperation between data science and environmental science in order to obtain and publicize results with certain scientific weight. Better and more improved data science technologies can help with that, so this kind of surveys can be taken on another level where all results will have a bigger impact on the subject.

5. CONCLUSIONS

Environmental science, along with physics, chemistry, biology, geosciences, biophysics, physical chemistry, etc., belongs to the basic natural science sector [13]. What makes it even more complex is the fact that it is a multidisciplinary science, which brings a certain level of responsibility to properly answer most of the miscellaneous questions. To answer those questions, researchers need to use a wide spectrum of analytical instruments. Those instruments can include processes such as data collecting, processing, and analysis, management and diagnostics.

Furthermore, it can be concluded that modern ecological reality requires a high level of understanding of the systems of the environment, which brings us back to the multidisciplinary approach for this branch of science [14].

The great challenge still remains to be the management of complex data, including interaction, crossing and using data from many different systems and fields of study.

The integration of data from environmental science is exceptionally important because we have a diversity of inputs from different sources. A new type of science emerged, which will have new approaches to overcoming those challenges [15].

The road towards a new type of science demands new incentives, which can be important for developing in a field of research. That included adaptive approaches in sampling and data collection. Furthermore, it will include adaptive modelling along with new techniques that are going to correspond with more complex and diverse environmental data, including data streaming in real-time [16].

Above all, the integration of qualitative data analysis into an environmental quality monitoring system is very important for assessing the socio-ecological status of a particular area.

To strengthen the symbiotic connection between data science and socio-environmental science takes time, however, that relationship could be a basis for further improvement of environmental policy making process.

REFERENCES

[1] D.A. Reed and J. Dongarra, "Exascale computing and big data", Commun. ACM 58, pp.56–68. doi: 10.1145 / 2699414, (2015).

[2] M. Potschin, R. Haines-Ioung, R. Fish, and R. Kerri Turner, "Routledge Handbook of Ecosystem Services", Nev Iork, NI: Routledge, (2016).

[3] F. Muller, R. de Groot, and L. Villemen, "Ecosystem Services at the Landscape Scale: the Need for Integrative Approaches", doi: 10.3097 / PO.201023, (2010).

[4] T. Hey, S. Tansley, and K. Troll, (eds.). "The Fourth Paradigm: Data-Intensive Scientific Discovery", Microsoft Research, (2009).

[5] C.L. Philip Chen, and C.I. Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data", Inf. Sci . 275, 314–347. doi: 10.1016 / j.ins.2014.01.015, (2014).

[6] C.S. Greene, J. Jie Tan, M. Ung, J.H. Moore, and C. Cheng, "Big Data Bioinformatics", J. Cell. Phisiol .229, 1896–1900. doi: 10.1002 / jcp.24662, (2014).

[7] H.V. Jagadish, J. Gehrke, A. Labrinidis, Y. Papakonstantinou, J.M. Patel, and R. Ramakrishnan, "Big data and its technical challenges". Commun. ACM 57, 86–94. doi: 10.1145/2611567, et al. (2014).

[8] E.S. Langley, A.A. Leeson, C.R. Stokes, and S.S.R. Jamieson, "Seasonal evolution of supraglacial lakes on an East Antarctic outley glacier", Geophys. Res. Lett. 43, 8563–8571. doi: 10.1002/2016GL069511, (2016).

[9] V. Nundloll, B. Porter, G.S. Blair, B. Emmett, J. Cosby, and D. Jones, "The design and deployment of an end-to-end IoT infrastructure for the natural environment", Future Intern. 11:129. doi: 10.3390/fi11060129, et al. (2019).

[10] M.A. Godard, A.J. Dougill, and T.G. Benton, "Scaling up from gardens: biodiversity conservation in urban environments", Trends Ecol. Evol. 25, 90–98. doi: 0.1016/j.tree.2009.07.016, (2010).

[11] D. Inđić, M. Terzić, and M. Ivanković, "Informacioni sistemi u upravljanju rizicima u životnoj sredini", Vojnotehnički glasnik, Vol. LXI. No. 1, pp. 210-225, DOI: 10.5937/vojtehg61-1678, (2013).

[12] M.R. Adžemović, A.Z. Šašić Ivanović, D.M. Nikolić, J.D. Aleksić, and R.M. Stevanović, "Redesigning current instruments as a precondition for a more effective environmental protection", Chemical Industry,71(4):343-350.DOI: 10.22987HEMIND160620045A, (2017).

[13] E.D. Ender and B.F. Smith, "Environmental Science – A Study of Interrelationships", 11th edition, Mc Graw-Hill International Edition, New York, (2008).

[14] A. Mihajlov, A. Jovanović, F. Jovanović and Hristina Stevanović-Čarapina, "Reporting by selected Analytical Tools for Environmental Management: Hazardous waste management in South Eastern Europe" , Journal of Environmental Protection and Ecology, 12(2), pp. 565-569, (2009).

[15] S.L. Cornford, D.F. Martin, D.T. Graves, D.F. Ranken, A.M. Le Brocq, R.M. Gladstone, and W.H. Lipscomb, "Adaptive mesh, finite volume modeling of marine ice sheets", J. Comput. Phys. 232, 529–549. doi: 10.1016/j.jcp.2012.08.037, et al. (2013).

[16] G.S. Blair, P. Henrys, A. Leeson, J. Watkins, E. Eastoe, S. Jarvis, and P.J. Young, "Data Science of the Natural Environment: A Research Roadmap", Front. Environ. Sci. 7:121. doi: 10.3389/fenvs.2019.00121, (2019).