EAI Endorsed Transactions
on Context-aware Systems and Applications Research Article EAI.EU

Open Data for Environment Sensing: Crowdsourcing Geolocation Data

Ngoan Thanh Trieu[1], Zachary E. S. Williams[2], Jean-François M. Dorville[3], Hiep Xuan Huynh[1,*], Vincent Rodin[4], Bernard Pottier[4]

[1] Can Tho University, Can Tho, Vietnam
[2] University of The West Indies, Kingston, Jamaica
[3] The Caribbean Geophysical and Numerical Research Group, Baie-Mahault, Guadeloupe
[4] Université de Bretagne Occidentale, Brest, France

Abstract

There are numerous situations where the digital representation of the environment appears critical for understanding and decision-making: threats on soils, water, seashores, risk of fires, pollutions are evident applications. If spatial cellular decomposition is evidence in the more common applications, there remains a large field for environment and activities modelling. The integration and composition of several information sources is perhaps the main difficulty with the need to deal with data interpretation and semantics inside concurrent simulators. Besides, the data on population, people's behaviours, people's perceptions are essential in environmental assessments, where the technical aspect is not counted as much as the common acceptance of impact technology. We provide a model for building environmental services with open data systems. A case study is given for getting information from the public about their relationship with freshwater and its scarcity in Jamaica.

Keywords: Open Data, Web Semantic, Environment Sensing, Geolocation Data, And Environmental Simulation.

Received on 29 February 2020, accepted on 09 May 2020, published on 12 May 2020

doi: 10.4108/_____

*Corresponding author. Email: hxhiep@ctu.edu.vn

1. Introduction

With the Internet development, the connection between people is better supported. The data generated from this connection can be up to the volume of exabytes or even zettabytes. Many challenges arise with this large amount of data such as data storage, processing, and leveraging the value of the data [1]. The data can include geographic information, environmental information, public health, education, statistics, etc. These data are stored under different formats and are kept in separate storage of organizations. There are almost no links between these data that allow the aggregation of different data sources.

Open data is referred to as a solution to this problem. Open data [2, 3] is the data that anyone can use and redistribute. The most recent research [9] points out the usefulness of combining crowdsourcing [26] (a large number of users in data creation) and sensing for a smart city. Data is collected from sensors, bus operating companies, and users to provide complete paths information according to individual's needs. This research provides applications that support people moving in Smart City by equipping them with accessible and personalized paths.

The urbanization process is causing negative impacts on the environment [4], contributing to pollution and harming human health. In this paper, we will give a model for building an open data system of environment sensing. The aim is to have a full path from collecting environmental data, creating open databases, and using the data for environmental simulations to provide warnings for specific issues. We give a case study that has

goal of information and data transparency commitments of many countries.

The main principles when considering open data are:

- Accessibility: Data must be available to a wide range of users and a variety of purposes. Protocols and formats of data delivery must be standard.
- Processability: The data provided must be organized so that it is convenient for automatic processing. The usability of the data is influenced by properly encoding the data.
- Globality: People must be able to use data without distinction between groups or domains.

The OpenSense Project [8] aims to provide the most convenient and efficient mechanism for monitoring air pollution. This is an important issue because it directly affects human health, especially in big cities where air pollution is getting worse. This project attaches sensors on public transport systems to collect data everywhere quickly and reduce the cost of installing sensors in multiple places. The large-scale environmental monitoring has posed many challenges for real-time handling of large data.

Open data is often associated with crowdsourcing data production [26], which means the involvement of a large number of users in data creation. With the participation of many users, the tasks will be done quickly and at a lower cost. An example is Wikipedia[1], an international online project for creating a free encyclopedia in multiple languages. Another example similar to Wikipedia is OpenStreetMap[2], the goal is to create a set of map data to freely use and edit. Users can download portions of OpenStreetMap information in vector or raster formats for later processing.

2.3 Linked Open Data

Linked data [5][6] is an important term in the concept of the Semantic Web. It means to create databases that can be understood by human and machine. In other words, this is the creation of a set of design principles for sharing machine-readable linked data on the Web. Machine-readable data [7] can be RDF, XML, and JSON.

Tim Berners-Lee outlines the five-star principles of Linked Data:

- Making data available on the Web
- Making data available as structured data
- Making data in a non-proprietary format
- Use URI to identify things, so that people can point at the data
- Link the data to other data to provide context

The Linking Open Data project developed by the W3C community[3] has put a lot of effort to enrich the linked open data cloud. This project has published various open datasets (such as DBPedia[4], Musicbrainz[5], DBLP[6], and Geonames[7]) as RDF on the Web. By interlinking, the user can navigate between DBPedia data to extra information provided by many different sources. Data is interconnected on a large scale allowing users to get more useful information from external databases when developing applications.

3. Environmental Simulation

3.1 Real-time monitoring

Many changes are appearing in climate, life, and economy balances. Fortunately, scientific activities brought knowledge and methods that give the hope to find solutions to rising problems. Domains such as meteorology, atmosphere studies, oceanography, agriculture, and biology are efficient and sometimes well organized.

It is known that some changes are very difficult to measure and monitor. Biodiversity and density of species are examples of the difficulties rising for measuring wide and sparse phenomena. Mekong Delta is infested by billions of insects that can destroy rice production and water salinity is invading the land putting even more pressure on agriculture. But there is no immediate way to classify and count insects, and for the physical underground water penetration, it is the same.

The core of research-oriented to climate change needs elaborated tools and techniques to collect physical information, to process this information and synthesize scientific facts accurately. Sensing is one part of the problem and deduction of distributed behaviour from local measures is another part. From an understanding of a physical, biological, or social status, it becomes an obvious issue to deduce possible evolutions and the effectiveness of counter-actions.

Previous research efforts associating these aspects can be mentioned for insect monitoring [21], building context-aware communication systems, and simulating physical phenomena [20]. From an understanding of physical, biological or social status, it becomes an obvious issue to deduce possible evolutions and the effectiveness of counter-actions. These efforts are currently improved using highly parallel computations [22] over a wide area and fine resolutions.

3.2 Environmental simulation

[1]https://www.wikipedia.org/

[2]https://www.openstreetmap.org

[3]https://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData

[4]https://wiki.dbpedia.org

[5]https://musicbrainz.org

[6]https://dblp.org

[7]https://www.geonames.org

The methodology is based on a cellular decomposition of geography. Practically, cells will embed information extracted from a database, completed by other geo-localized data coming from different sources. It is currently the case for elevations used to model radio signal propagation or rain flooding simulation. It will be the case for other information coming from sensor fields, satellite image analysis, and feedback information from the public.

Current tools are presented in [20], they address geographic position, sensor network abstraction, and physical representation based on cell systems. The tools enable fast production of high-performance simulators yet ready for concurrent process networks, and graphic processing units, and soon supercomputing with scales of millions of cells and hundred of squared kilometres. The systems are animated using a computing method called "Cellular Automata". We will keep these core functionalities, opening the input data integration, and producing result publications as web services.

The current development efforts[8] include:

- Database storage based on Postgis support, and OpenStreetMap
- Serving tiles for local (Quickmap) and remote browsers (OpenLayers)
- Generation of high-performance concurrent simulators (Multicores, GPUs, MPI)
- Service software architecture for remote end-users (Seaside).
- External data integration in database: meteorological radar map, elevations, sensor fields

The core objective is environment sensing and simulation, in evolving aspects and larger information fields. This includes the support for open data integration, production and publication of predictions coming from simulations, direct interaction with engineers, specialists, and in some place interested publics.

We propose a model for building environmental services with open data (figure 4). The whole process is combining of environmental data collection, open databases creation, and environmental services formation.

4. Case study: Crowdsourcing geolocation data

How, in a real case, useful data can be generated, linked, and made accessible to the worldwide population. This section presents an example of crowdsourcing data collection, a new trend of data collection with help of a large group of people.

4.1 Environmental context and needs of data

Freshwater is a vital element for human beings but since the industrial revolution is becoming one of the most endangered resources [11]. Population growth but also the increase of the need in agriculture and industry in a climate change context-induced more scarcity of the 3% freshwater available at earth's surface. Solutions are available to face these new needs of freshwater particularly during the dry season and drought events. Desalination is based on the usage of the main water resource available at the earth's surface (i.e. salted water) to produce drinkable water [24]. Water molecules are segregated from dissolved salt by thermal, chemical or mechanical methods. Most of the eight main methods of desalination use large amounts of energy to produce freshwater and a highly salted waste. The brine produced can be converted in salt but is more frequently released at the coastline area with negative effects on the environment [10].

Quality and taste of the freshwater produced by the desalination process are not the same as springs, rivers, or well water. A shift from conventional freshwater procurement to desalination cannot be done without a large amount of energy, full access to high seawater quality, and a population ready to change its water usage habits. That change involves data to design the industrial plant, determine its best location, ease water resource management, and evaluate how people will be able to adapt or not.

As part of a research project to design a desalination plant powered by mix renewable energy i.e. wind-solar-

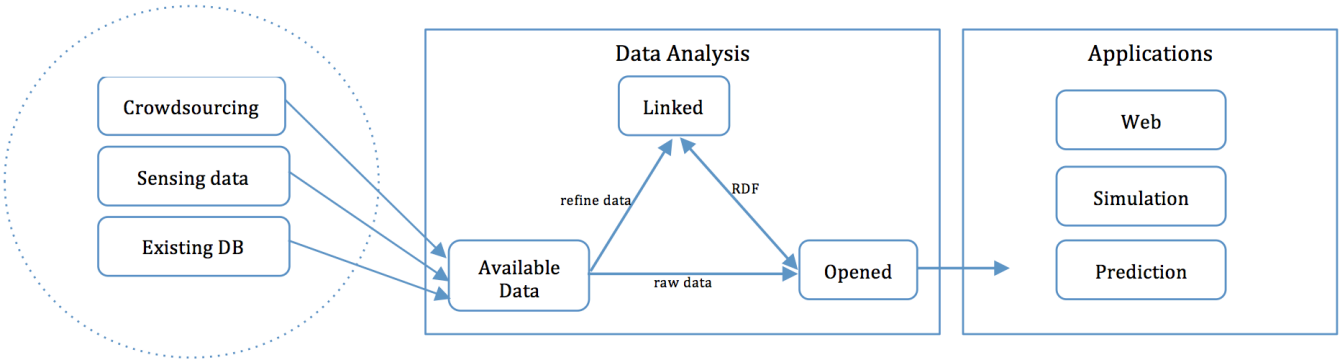

Figure 4. Model for Environmental services with Open

[8]http://sames.univ-brest.fr/sameswp/

wave for the island of Jamaica (\#JamGeenDesal), the needs of appropriate data was highlighted. Official open

4.5 Generation of open data

The structure of the collection process can push to use a relational database where each question class is stored in a dedicated table or sub-table related by an index. Access to the database can be a limit to the concept of Open Data.

Another organization of the results can be chosen to ease the dissemination of the information, it is base on RDF format [23] through a dedicated XML file or web page using RDF format (Figure 5). This web page or XML file will summarize the results in human-readable format, with information linked to main features of the participants i.e. age, gender, location, profession and interest to environmental issues.

4.6 Perspectives

A second phase of the process will be launched soon with a large target audience with a better selection of the questions and modes of answers. The second phase will include a Geographic Information Systems (GIS) tool to get information with lower space units: at the scale of a city or of a district or even smaller. That small unit size will be close to the cells used for environmental simulations and ease the integration of human behaviour in modelling. It will also take into account time and integrate meteorological seasons in questions/answers. Data analysis will be designed to be used as fully open and linked data.

5. Conclusion

This study reveals the major interests of environmental sensing and simulation in prediction physical issues. A clear model is given showing how we collect environmental information and create open data for building environmental services. Cellular automata with transition rules between cells are the core concept in this work for simulations. Open data is hoped to give the vision of ambitious information systems covering the environment in the neighbourhood.

Acknowledgements.

This work is supported by a Brest Metropole initiative related to Open Data for environment. It is rooted in SAMES project (Ministère des Affaires Etrangeres 2016-2018) grouping mainly researchers from UBO, Can Tho University, ... and now from places in West Indies, Asia, and Africa around software developments for environment modelling and simulation.

References

[1] Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. Big data: Issues and challenges moving forward. *46th Hawaii International Conference on System Sciences.* January 2013. IEEE. p. 995-1004.

[2] Evans, J. A., & Reimer, J. Open access and global participation in science. *Science.* 2009; *323*(5917): 1025-1025.

[3] Xu, G. H. Open access to scientific data: promoting science and innovation. *Data Science Journal. 2007; 6*: OD21-OD25.

[4] Uttara, S., Bhuvandas, N., & Aggarwal, V. Impacts of urbanization on environment. *International Journal of Research in Engineering and Applied Sciences*. 2012; *2*(2): 1637-1645.

[5] Bizer, C., Heath, T., & Berners-Lee, T. *Semantic Services, Interoperability and Web Applications: Emerging Concepts.* IGI Global; 2011. Linked data: The story so far; p. [205-227].

[6] Bizer, C., Heath, T., Idehen, K., & Berners-Lee, T. Linked data on the web (LDOW2008). *Proceedings of the 17th International Conference on World Wide Web.* April 2008. p. 1265-1266.

[7] Barometer, O. D. Open data barometer. *World Wide Web Foundation.* 2015; 1-60.

[8] Aberer, K., Sathe, S., Chakraborty, D., Martinoli, A., Barrenetxea, G., Faltings, B., & Thiele, L. OpenSense: Open community driven sensing of environment. *Proceedings of the ACM SIGSPATIAL International Workshop on GeoStreaming.* November 2010. p. 39-42.

[9] Mirri, S., Prandi, C., Salomoni, P., Callegati, F., & Campi, A. On combining crowdsourcing, sensing and open data for an accessible smart city. *Eighth International Conference on Next Generation Mobile Apps, Services and Technologies.* IEEE; September 2014. p. 294-299.

[10] Jones, E., Qadir, M., van Vliet, M. T., Smakhtin, V., & Kang, S. M. The state of desalination and brine production: A global outlook. *Science of the Total Environment*. 2019; *657*: 1343-1356.

[11] Oki, T., & Kanae, S. Global hydrological cycles and world water resources. *Science*. 2006; *313*(5790): 1068-1072.

[12] Roßmann, J., Gummer, T., & Silber, H. Mitigating satisficing in cognitively demanding grid questions: evidence from two web-based experiments. *Journal of Survey Statistics and Methodology*. 2018; *6*(3): 376-400.

[13] Z. E. S Williams. Renewable Energy for Desalination Process: Efficiency and Environmental Impacts in a Tropical Island Using Digital Tools. Master thesis, Jamaica*: Univ. The West Indies*; 2019.

[14] Berners-Lee, T., Hendler, J., & Lassila, O. The semantic web. *Scientific American*. 2001; *284*(5): 34-43.

[15] Euzenat, J., & Shvaiko, P. *Ontology Matching*. Volume 18. Heidelberg, Berlin: Springer; 2007.

[16] McGuinness, D. L., & Van Harmelen, F. OWL web ontology language overview. *W3C Recommendation*. 2004; *10*(10): 2004.

[17] Antoniou, G., & Van Harmelen, F. *Handbook on Ontologies*. Heidelberg, Berlin: Springer; 2004. Web ontology language: Owl; p. [67-92].

[18] Miller, E. An introduction to the resource description framework. *Bulletin of the American Society for Information Science and Technology*. 1998; *25*(1): 15-19.

[19] Schmidt, M., Meier, M., & Lausen, G. Foundations of SPARQL query optimization. *Proceedings of the 13th International Conference on Database Theory.* March 2010. p. 4-33.

[20] Truong, T. P., Pottier, B., & Huynh, H. X. Cellular Simulation for Distributed Sensing over Complex Terrains. *Sensors*. 2018; *18*(7): 2323.

[21] Lam, B. H., Huynh, H. X., & Pottier, B. Synchronous networks for bioenvironmental surveillance based on cellular automata. *EAI Endorsed Transactions on Context-Aware Systems and Applications*. 2016; *3*(8).

[22] Truong, M. T. T., Samar, S. Y., Pottier, B., Rodin, V., & Huynh, X. H. Multiscale Geographic Exploration, Observation, Simulation, and Representation. *13th International Conference-Mathematics, Actuarial, Computer Science & Statistics (MACS 13)*. IEEE; December 2019. p. 1-8.
[23] Coy, C. et al. *The Jamaica Labour Force 2017: annual review.* Kingston, Jamaica: The Statistic Institute Of Jamaica: 2017.
[24] Buros, O. K. *The ABCs of Desalting*. Topsfield, MA: International Desalination Association; 2000.
[25] Shaffer, J. P. Multiple hypothesis testing. *Annual Review of Psychology*. 1995; *46*(1): 561-584.
[26] Howe Jeff. The rise of crowdsourcing. *Wired magazine.* 2006; 14(6): 1-4.