Research Challenges of Open Data as a Service for Smart Cities

Leonard Walletzký[1], Františka Romanovská[1], Angeliki Maria Toli[3] and Mouzhi Ge[1,2]
[1]*Faculty of Informatics, Masaryk University, Brno, Czech Republic*
[2]*Institute of Computer Science, Masaryk University, Brno, Czech Republic*
[3]*Bartlett School of Construction and Project Management, University College London, London, U.K.*

Keywords: Open Data, Data Life Cycle, Open Data as a Service.

Abstract: The open data are considered to be an important building block of Smart City services. Based on the data availability in the cities, companies are building innovative smart services to facilitate the city development. However, most of the practitioners are focusing only on the open data usage but paying little attention to the processes related to the data publication. Thus, there is a lack of understanding the whole process of open data life cycle such as before the data can be open, when they are published and when they need to be archived. Those phases in the open data life cycle are critical for assuring the data accuracy, availability and relevance because the data are analysed, changed, anatomized or generally processed during the whole life cycle. This also creates a set of research challenges for open data. Therefore, in the context of smart cities, this paper proposes to consider Open Data as a Service and identifies the research challenges along with the open data life cycle.

1 INTRODUCTION

The term smart city gained popularity in the early 2010s to demonstrate how technological advances along with data that have the significant potential to improve city planning and management (Stepánek et al., 2017). There is a variety of data sources that can be used in smart city services. For example, smart city projects regularly collect data via Internet of Things (IoT) platforms, where the collected data can further be published as open data. Thus, while the development of smart cities increases the open data, open data also facilitate the smart city applications. In smart cities, open data can be defined as data that are available and accessible for use and reproduction by any party (Lindman et al., 2013). Those open data can bring various advantages to smart cities, for example, governmental open data will promote transparency and accountability (McDermott, 2010), improve the decision making capacity and create innovation opportunities for both governmental and private actors. Moreover, open data can also improve the provision of public services and lead to the improvement of citizens' quality of life (Pereira et al., 2017). Due to high utility of open data, open data initiatives in smart cities are considered to have a significant impact on the domains of governance, economy and transport and mobility (Ojo et al., 2015b). However, literature suggests that the actual effect of open data on ongoing smart city projects has been elusive to assess. For example, the smart city program of Rio de Janeiro has been slow in achieving its goals regarding the open data (Angelidou, 2017).

Open data can help to increase the positive image and motivate the creation of the innovative applications. For example, in the Brno city (second largest city in Czech Republic with 380.000 inhabitants), open data show that total length of the roads of bikers was around 63km in 2018 and Brno is announcing that is having about 136km of the roads for bikers in 2019 open city data. This helps the city to build an image that the city is planning more cycling lanes and thus more and more applications are developed to support bikers in Brno in 2020. From this real-world scenario, it can be seen that open data can help citizens and organizations to understand the city and build the smart city together. With the increasing importance of open data, most organizations are considering the open data only as a data source and usually ignore how the life cycle of the open data and how to help to improve the healthy life cycle of open data. Therefore, this paper views the open data as a service and discusses how to create Open Data as a Service in smart cities.

Walletzký, L., Romanovská, F., Toli, A. and Ge, M.
Research Challenges of Open Data as a Service for Smart Cities.
DOI: 10.5220/0009578604680472
In *Proceedings of the 10th International Conference on Cloud Computing and Services Science (CLOSER 2020)*, pages 468-472
ISBN: 978-989-758-424-4

The rest of the paper is organized as follows, Section 2 reviews the open data and its life cycle. Then from the service perspective, Section 3 discusses the open data process, related service and value propositions. Section 4 proposes a set of research challenges of Open Data as a Service for Smart Cities, finally, section 5 concludes the paper and outlines future works.

2 SERVICE VIEW TO OPEN DATA

In order to understand the relations between smart cities and open data, the British Standards Institution envisions smart city as a city that is citizen-centred, collaborative, digital and defined by open data (BSI, 2014). Open data can be defined as data and content that can be used, modified and shared with no restrictions, by anyone and for any scope (ODI, 2017). Data can be characterized as open when they meet the following criteria: accessible -thus typically published on the web- , provided in a machine-readable format and under a license that allows to access, use and share them (Transport Systems Catapult, 2017). Open data initiatives support the smart city objectives and are considered to have a significant impact on the domains of governance, economy and transport and mobility (Ojo et al., 2015a). They can be viewed as a way to restore the separation between government and civic society by stimulating public organizations to act as a more open system (Janssen et al., 2012). Smart city projects regularly collect data via IoT platforms, which are frequently made open (Ahlgren et al., 2016). These data are often sourced by governmental bodies from citizens through sensors, kiosks, meters, smart phones and smart appliances (Harrison et al., 2010) and are made available by numerous cities through open data portals, based on the internet and connected to data hubs (Caird and Hallett, 2019). An example of such cities is London, UK. London offers a number of open data sources such as the London Datastore, Transport for London (TfL) and data.gov.uk, the main themes of which are: demographics, employment and skills, environment, transparency, housing, health, transport, business and economy and education (Ghahremanlou et al., 2019).

Literature suggests that open data operate in smart city ecosystems (Hall et al., 2012; Poikola et al., 2011; Zuiderwijk et al., 2014), characterized by interdependencies between issues (Zuiderwijk et al., 2014). Such issues may be related to pre-existing networks, power and information asymmetries and variance of training and capacity (Hall et al., 2012). (Zuiderwijk et al., 2014) present an open government data scenario to demonstrate how open data ecosystems operate, which correlates to data lifecycle management. This demonstrates how governmental organizations can create or collect, subsequently store and curate, then verify and remove sensitive information from the data that are then made available to users. These data are frequently made available, via data integrating actors referred to as infomediaries. Subsequently, users can return to the governmental organizations with further data requests or to discuss the datasets.

According to (Zuiderwijk et al., 2012) the process of opening the data can be split up into five steps. The first two steps are focusing on the responsibilities of a city/organization, such as creating the data and opening the data. The next two steps focus on users (finding the open data and using them) and the last step closes the cycle on a border between user and organization in which they discuss and provide feedback on the open data (see Figure 1). This describes the open data as an ongoing process with a constant need for improvement of the service they provide for the users; however, it does not discuss closing the open data. By that is meant the withdrawal of the published data due to various reasons, e.g., change in legislation. Only one (Demchenko et al., 2013) of the ten models describing the data life cycle in (Charalabidis et al., 2018) mentions a step after sharing/publishing the data.

3 OPEN DATA LIFE CYCLE AND VALUE PROPOSITION

From the perspective of Open Data as a Service, we propose to simplify the open data life cycle into three stages: before open data, open data and after open data. although it is important to describe the open data in each of the three stages, we found that most literature is focused on the open data application and lacking to pay attention to *before open data* and *after open data* stage.

From the perspective of Service dominant logic (Lusch and Vargo, 2014), we conduct our analysis from product view to service view in order to identify the value of the data and their usefulness to their users. The important characteristic of open data service is the value proposition. The value proposition refers to the potential value the customer can obtain by usage of particular service. Therefore, considering the three basic steps of open data life cycle, we can investigate the value proposition that is designed for Open Data as a Service.

Before Open Data: value proposition is based on

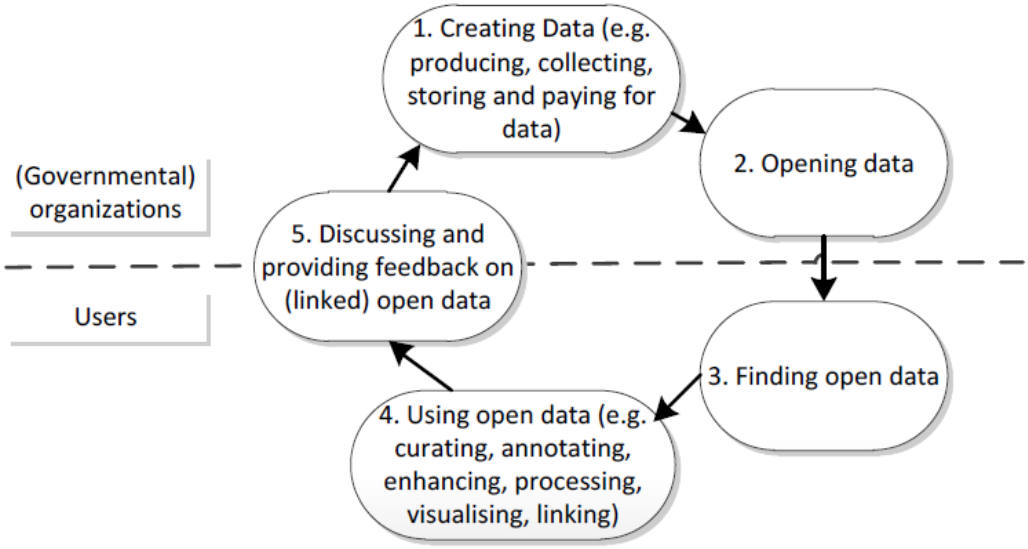

Figure 1: Open data interactions between organizations and users (Zuiderwijk et al., 2012).

the possibility and ability of the data to be transformed into Open Data.

Open Data: value proposition is focused on the ability of data to create value, for example, to be an input to some innovative application or provide relevant information to stakeholders.

After Open Data: value proposition is about saving historical values and being able to provide retrospective analyses.

It can be seen that each part of open data life cycle has different value proposition, also focused to the different stakeholders. Therefore, the service of open data, its features and aspects, must be analysed separately, because the consequences and results are different in each phase of life cycle.

4 OPEN DATA AS A SERVICE

According to (Zuiderwijk et al., 2014), the following aspects of open data has been identified:

- Availability and Access: collecting data; deciding what to collect, when and by whom it should be collected; not opening the data
- Find Ability: the ability of data and metadata to be easily found by users, as well as easily browsed and searched
- Usability: Ensuring the correctness of the data and trustworthiness of the source. The right amount of services to use the data.
- Understand Ability: The ability of the data to be understood by users (e.g., visualisation) supported by good and user friendly API. Knowledge of users to use the data.
- Quality: Reliable and accurate data.
- Linking and Combining Data: The ability of the data to be linked and combined.
- Comparability and Compatibility: Unified and standardized format of data for easier comparison. Uniform policy for publishing.
- Metadata: Commonly agreed metadata with recognizable structure, their extent and provision.

From our applied research results, based on cooperation with public authorities and city representatives, we identified three critical aspects:

Security and Privacy. This aspects is mostly underestimated by municipalities and stakeholders. This is a very complex problem, combining issues of data privacy, authorization and availability. As the data are public available, the data privacy and data anonymization become very important to protect the citizen's personal information (Mbarek et al., 2020).

Quality. Data quality research can be traced back to two decades ago (Ge and Helfert, 2006; Helfert et al., 2009b), in the context of open data, while the open data facilitate to share the information and knowledge among the Internet peers, it also suffers from a large variety of the data quality problems (Helfert et al., 2009a; Ge et al., 2011). Consider the Wikipedia as an example, (Lewoniewski et al., 2017) report that for the same fact number such population of certain city, different language versions of the same Wikipedia articles can be different. Further, the contents for the same topic in different language versions also vary significantly.

Quantity or Volume. Not only data quality is important, but also data quantity plays a critical role. For the most of the people from municipalities, open data are represented only in the form of tables (more specifically in csv format). However there are more concerns, for example, we can have a data streams where only part of the data is relevant, or we can have the Big Data, where only some of data are important for decision-making (Ge et al., 2018). To further refine this research issue, it is also hard to identify which part of Big Data is important for a specific application or use case. Compared to other Vs of Big Data, the volume of big open data should be prioritized.

It can be seen that each aspect can play an important role in the life cycle of the open data. Along the open data life cycle, data access and usability may vary. Before open data, the critical aspect is to create and process open data, we need to assure to obtain the right data in the right time. When the data is opened, it is important to concern the usability for the city and third-party organizations. It is representing the feature to be used to create a value for the users of applications, linked to open data. After opening the data, the related questions can be that what data can be stored and what data can be deleted? If they are stored, for how long they should be available?

We found that the research of open data for smart cities can be done in the cooperation with municipalities in the different stages of Open Data development. If the research would be done in the municipality of the same level, some of critical aspects of problems can be missed or ignored. If this analysis can be done, it will clarify the process of Open Data publication, help the developers of Open Data platforms to design their products more precisely and feed the data service need of all relevant stakeholders during whole data life cycle.

5 CONCLUSION

In this position paper, we have proposed to underpin the importance of the open data life cycle and consider open data as a service. We have revisited how the open data are generated and used along its life cycle in smart cities. From the service perspective, we have investigated the value proposition of the open data in the phase of before open data, open data and after open data. Based on studying the open data from service perspective, we have reviewed and identified as a set of research issues for open data as a service, where we especially propose to focus on three critical aspects for open data in smart cities: security, quality and quantity. Each aspects have been discussed to facilitate the future research of open data in smart cities, and also help to develop the concept of Open Data as a Service for smart cities.

ACKNOWLEDGEMENTS

The work is supported from European Regional Development Fund Project *CERIT Scientific Cloud* (No. CZ.02.1.01/0.0/0.0/16_013/0001802).

REFERENCES

Ahlgren, B., Hidell, M., and Ngai, E. (2016). Internet of things for smart cities: Interoperability and open data. *IEEE Internet Computing, IEEE*, 20(6):52–56.

Angelidou, M. (2017). Angelidou, m. (2017), "smart city planning and development shortcomings", tema. journal of land use, mobility and environment, vol. 10 no. 1, pp. 77-94. *TeMA - Journal of Land Use, Mobility and Environment*, 10:77–93.

BSI (2014). *PAS 181:2014 Smart City Framework – Guide to Establishing Strategies for Smart Cities and Communities*. BSI Group.

Caird, S. and Hallett, S. (2019). Towards evaluation design for smart city development. *Journal of Urban Design, Routledge*, 24(2):188–209.

Charalabidis, Y., Zuiderwijk, A., Alexopoulos, C., Janssen, M., Lampoltshammer, T., and Ferro, E. (2018). The multiple life cycles of open data creation and use. In Charalabidis, Y., Zuiderwijk, A., Alexopoulos, C., Janssen, M., Lampoltshammer, T., and Ferro, E., editors, *The World of Open Data*, page 20 – 21. Springer International Publishing. Online]. Available:.

Demchenko, Y., Grosso, P., De Laat, C., and Membrey, P. (2013). Addressing big data issues in scientific data infrastructure. In *Proceedings of the 2013 International Conference on Collaboration Technologies and Systems, CTS 2013*.

Ge, M., Bangui, H., and Buhnova, B. (2018). Big data for internet of things: A survey. *Future Gener. Comput. Syst.*, 87:601–614.

Ge, M. and Helfert, M. (2006). A framework to assess decision quality using information quality dimensions. In *Proceedings of the 11th International Conference on Information Quality, MIT, Cambridge, MA, USA, November 10-12, 2006*, pages 455–466.

Ge, M., Helfert, M., and Jannach, D. (2011). Information quality assessment: validating measurement dimensions and processes. In *19th European Conference on Information Systems, ECIS 2011, Helsinki, Finland, June 9-11, 2011*, page 75.

Ghahremanlou, L., Tawil, H., Kearney, A., Nevisi, P., Zhao, H., X., and Abdallah, A. (2019). A survey of open data platforms in six uk smart city initiatives. *The Computer Journal*, 62(7):961–976.

Hall, W., Shadbolt, N., Tiropanis, T., O'Hara, K., and Davies, T. (2012). Open data and charities. a state of the art review.

Harrison, C., Eckman, B., Hamilton, R., Hartswick, P., Kalagnanam, J., Paraszczak, J., and Williams, P. (2010). Foundations for smarter cities". *IBM Journal of Research and Development*, 54(4):1–16.

Helfert, M., Foley, O., Ge, M., and Cappiello, C. (2009a). Analysing the effect of security on information quality dimensions. In *17th European Conference on Information Systems, ECIS 2009, Verona, Italy, 2009*, pages 2785–2797.

Helfert, M., Foley, O., Ge, M., and Cappiello, C. (2009b). Limitations of weighted sum measures for information quality. In *Proceedings of the 15th Americas Conference on Information Systems, AMCIS 2009, San Francisco, California, USA, August 6-9, 2009*, page 277.

Janssen, M., Charalabidis, Y., and Zuiderwijk, A. (2012). *Benefits, adoption barriers and myths of open data and open government*, volume 29. Information Systems Management, Taylor & Francis.

Lewoniewski, W., Węcel, K., and Abramowicz, W. (2017). Relative quality and popularity evaluation of multilingual wikipedia articles. *Informatics*, 4(4):43.

Lindman, J., Rossi, M., and Tuunainen, V. K. (2013). Open data services: Research agenda. In *46th Hawaii International Conference on System Sciences, HICSS 2013, Wailea, HI, USA, January 7-10, 2013*, pages 1239–1246.

Lusch, R. F. and Vargo, S. L. (2014). *The service-dominant logic of marketing: Dialog, debate, and directions*. Routledge.

Mbarek, B., Ge, M., and Pitner, T. (2020). An efficient mutual authentication scheme for internet of things. *Internet of Things*, 9:100160.

McDermott, P. (2010). Building open government. *Government Information Quarterly*, 27(4):401–413.

ODI (2017). *Using Open Data to Deliver Public Services*. Open Data Institute, London.

Ojo, A., Curry, E., and Zeleti, F. (2015a). A tale of open data innovations in five smart cities". In *Proceedings of the Annual Hawaii International Conference on System Sciences, IEEE*, page 2326–2335.

Ojo, A. K., Curry, E., and Zeleti, F. A. (2015b). A tale of open data innovations in five smart cities. In *48th Hawaii International Conference on System Sciences, HICSS 2015, Kauai, Hawaii, USA, January 5-8, 2015*, pages 2326–2335.

Pereira, G. V., Macadar, M. A., Luciano, E. M., and Testa, M. G. (2017). Delivering public value through open government data initiatives in a smart city context. *Information Systems Frontiers*, 19(2):213–229.

Poikola, A., Kola, P., and Hintikka, K. (2011). *Public data—an introduction to opening information resources, Ministry of Transport and Communications*. Http://Www. Scribd. Com/Doc/57392397/Public-Data, Helsinki.

Stepánek, P., Ge, M., and Walletzký, L. (2017). It-enabled digital service design principles - lessons learned from digital cities. In *Information Systems - 14th European, Mediterranean, and Middle Eastern Conference, EMCIS 2017, Coimbra, Portugal, September 7-8, 2017, Proceedings*, pages 186–196.

Transport Systems Catapult (2017). In *THE CASE FOR GOVERNMENT INVOLVEMENT TO INCENTIVISE DATA SHARING IN THE UK INTELLIGENT MOBILITY SECTOR*.

Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R., and Alibaks, R. (2012). Socio-technical impediments of open data. *Electronic Journal of Electronic Government*, 10(2):156–172.

Zuiderwijk, A., Janssen, M., and Davis, C. (2014). Innovation with open data: Essential elements of open data ecosystems. *Information Polity*, 19(1–2):17–33.