From Access to Effective Use: Open Data Portals for Everyday Citizens

Unisse C. Chua
De La Salle University
Manila, Philippines
unisse.chua@dlsu.edu.ph

Kyle L. Santiago
De La Salle University
Manila, Philippines
kyle_santiago@dlsu.edu.ph

Ian Benedict M. Ona
De La Salle University
Manila, Philippines
ian_ona@dlsu.edu.ph

Romeo Manuel N. Peña
De La Salle University
Manila, Philippines
romeo_pena@dlsu.edu.ph

Geremiah Zachary S. Marasigan
De La Salle University
Manila, Philippines
geremiah_marasigan@dlsu.edu.ph

Paolo Gabriel A. Delos Reyes
De La Salle University
Manila, Philippines
paolo_delosreyes@dlsu.edu.ph

Briane Paul V. Samson
Future University Hakodate
Hakodate, Japan
De La Salle University
Manila, Philippines
briane.samson@dlsu.edu.ph

ABSTRACT

Open government data allows for transparency from governments and access to data collected about its citizens. However, we are still far from achieving universal citizen participation because data literacy and experience are necessary to extract insights from data. There is also no guarantee if available data can address people's information needs. We explored the potential of open government data portals in addressing the information needs of citizens through an online survey and found that these can only be partially answered by the available data. To understand their information seeking behavior, we conducted usability tests of open data portals with 21 citizens, and used semi-structured interviews to identify gaps in the portals' design. We found that citizens would benefit from: localized and advanced search engines; and visualized, contextualized, and processed content for better sensemaking. We conclude with design guidelines for open data portals catered to citizens.

Author Keywords

Open data, information seeking, user study

CCS Concepts

•**Human-centered computing** → **Human computer interaction (HCI)**; *User studies;* Usability testing;

INTRODUCTION

Through the advancement of internet and web technologies, various governments have made their services and data publicly available by publishing open government data (OGD) [5, 17]. Government agencies, civic hackers and third-party intermediaries have been maximizing the use of open data for development and innovation. Several technology companies, such as Sakay.ph[1] and Rentlogic[2], leveraged on open data in developing solutions that provide commute routes in the Philippines and apartment building ratings in New York, respectively. Among those involved in open data projects, their common activities can be categorized into technical tasks such as converting data, programming, ensuring data quality and non-technical tasks such as requesting data, promoting [7]. This and other user studies have so far focused on technical users [3, 6, 8, 9, 10]. Although these portals and innovative solutions using open data have been perceived to promote citizen engagement [1], less is known about how everyday citizens seek and engage with open government data. Our pilot study focuses on understanding if and how everyday non-technical citizens address their information needs about the government and identifying design elements that would potentially aid in engaging them to maximize open government data.

---

[1] https://sakay.ph/

[2] https://www.rentlogic.com/

RELATED WORK

A survey of research in OGD implementation and human-data interaction (HDI) suggests the need for insights from lay citizens since they are also the target users of OGD portals [15]. Existing HCI research provides ways of improving OGD portals for experienced data users [3, 6, 8, 9, 10]. These users are usually from open data intermediaries (i.e. civic start-ups and infomediaries) that usually take the role of aggregators, developers, enrichers, and enablers [14]. However, little research has explored the potential of OGD portal use for lay citizens. Use of OGD by citizens may be compromised due to the restricting understandability of raw data [16]. Because of this, we aim to design a more inclusive OGD portal that caters to both types of users, along with proper policy and promotion, may increase citizen engagement in open government.

PHILIPPINE OPEN DATA SYSTEMS

As one of the pioneering members of the Open Government Partnership in 2011, the Philippines has taken steps to disclose data to the public through the Full Disclosure Policy [2]. Since then, the Philippines has launched various portals containing OGD [2, 4, 15]. In this study, we focus on two specific portals: Open Data Philippines (ODPH) and the Electronic Freedom of Information (eFOI). These portals cover the two ways citizens can gain access to OGD: government-led publication of data and citizen-led request of information. ODPH[3] hosts data provided by different agencies and allows users to search, preview, and download these without any restrictions. In the dataset preview, users may also create exploratory graphs with the fields available in the file. On the other hand, eFOI[4] serves as an online tracker of citizen requests for government information that has not yet been published. eFOI requires citizens to register and to provide a valid proof of identity before making a request. Requests must be filed to the correct agency otherwise it gets rejected. An email would be sent notifying the requester of the status after submission. Other citizens may also browse through existing requests. Each request shows a conversational thread of the FOI officer and the corresponding agency.

NEEDFINDING SURVEY AND PORTAL ANALYSIS

We designed our survey to be as extensive as possible to understand the factors that could drive the awareness and the usage of open government portals. We organized it in sections: Experience on Data Work, Information Seeking, Awareness and Usage of ODPH and eFOI.

We asked if they have done any data analysis work before answering the survey. To understand the information seeking behavior of Filipinos, we provided three scenarios with varying degrees of personal involvement in decision making based on a previous study highlighting how data is personal [12]. The personal scenario involves making a decision involving oneself. The extrapersonal scenarios may affect the individual to some extent but are more focused on another person, while the impersonal scenario involves looking for information that is more detached from their personal lives. For each

scenario, we asked the participants which information sources they would most likely consult to make the decision. Before asking how often they visited the portals and how useful they found the data within the portal, we provided them a brief overview of the portals. In the end, we asked them to provide three government sectors they would be most interested in using the categories present in ODPH with an "Others" option. The online survey was given in English and Filipino.

In addition to the survey, we also conducted an inventory analysis of ODPH. The open-ended response from the wild card scenario was analyzed and cross-checked with the available data from ODPH and eFOI to determine if these queries are answerable by existing data. This was done by one of the authors who has experience with data science projects.

Participants

We distributed the online survey to Filipino citizens through social networking sites and online communities. More than half (N=63) of our 119 respondents were women. The average age was 33 years old (SD=13.77).

Results

We were able to collect a total of 119 valid responses from both forms with six of them coming from the Filipino survey. Our results show that participants have varying sources of information depending on the degree of personal involvement of the situation. Family and friends are more likely sources for personal situations. In addition to family and friends, Google is an additional source for extrapersonal situations. For impersonal situations, news websites and Google search were the top choices. When asked if they had any other data-related inquiries, Google was the main information source for majority of the respondents (N=77; 64.71%). From these results, it shows that people have a high dependency on big search engines. In terms of awareness of the portals, we found that people were generally more aware of ODPH than eFOI. We found that the majority of citizens (N=39) were unaware, while only 11 were aware of both portals. For the usage frequency and usefulness of the data found in the portal, the overall response from the participants was that the data provided was *"either incomplete or non-existent"*. Looking into the wild card scenario provided by the participants, people were generally interested in potential housing locations (N=18), public transportation routes and improvements (N=15), and economic development (N=10). We classified these queries to be *"answerable"* based on existing knowledge on what kind of datasets may be used to provide either a partial or complete information. Based on this, we found that 24 out of 107 (22.43%) responses may potentially be answered in part by available data in ODPH.

USABILITY EXPERIMENTS

To further examine the information seeking behavior of citizens, we conducted experiments on how citizens use ODPH and eFOI. We also conducted semi-structured interviews in parallel to identify gaps in the design of the portals. Dialogues were audio recorded while the on-screen behavior was recorded using a screen capture software. There were two parts in the experiment: (1) an open search, wherein participants can search freely, and (2) separate tests for ODPH and

eFOI. In each part, we provided two tasks: personal tasks associated with one's place of residence, as people associate better with data when it is linked to a place [13], and non-personal tasks. Each participant was asked to perform the same tasks with the place of residence as the only variable in the personal task. All tasks were guaranteed to be answerable by available data in either portals. After each completed task, we follow up with post-task interview questions.

Participants

We recruited individuals with less than a year of experience with data analysis and no experience using OGD portals for any kind of work. We defined data analysis as the ability to handle raw data, process and visualize the data. Participants that meet either criteria were classified as a *lay person* (L) which represent ordinary citizens. Otherwise, we identified them as *data workers* (DW).We were able to successfully recruit 13 lay people and 8 data workers for our interviews. However, 2 of the lay people expressed their lack of experience in working with computers in general so we separated them to *computer illiterate* (SL).The average age for all participants was 31 (SD=15.08) and majority (N=14) were men.

Results

For our analysis, we excluded the 2 computer illiterate participants since they were both unable to complete any of the tasks given without any assistance from the researchers. For the open search tasks, all participants were able to complete at least one of the two tasks provided (avg. completion rate: DW=87.5%, L=77.3%). We observed that majority (N=16) of our participants found it challenging to find datasets in ODPH because of the **unoptimized search engine** (avg. completion rate: DW=81.3%, L=77.3%). Most of queries used would return no results and participants thought it would work like Google. We analyzed the search queries and the results echo the findings from a previous study [8] wherein spatial and temporal keywords are predominantly used to filter results depending on the task at hand but these were not integrated in the data search functions. We also noticed that data workers refined their search queries more using special keywords such as file formats and agencies responsible for the data. ODPH has **categorical icons** below the search bar which provides dataset filtering however, only seven participants used this feature. In the previous design, the icons were not visible on load which suggest that this feature may be underutilized. For eFOI, both groups had a harder time completing tasks (avg. completion rate: DW=75.0%, L=59.1%). Some participants (N=5) disliked the **conversational structure** and three participants specifically mentioned the lack of hyperlinks in the text which made it difficult to easily skip through the text and simply click the link to the data.

We also observed that lay participants preferred information presented in web articles with a mixture of numbers and text (45.2%) and tabulated data (38.7%). These participants preferred straightforward presentation of information with organization and labels. Data workers prefer raw datasets (24.1%) and tabulated data (41.4%). We believe that this emphasizes the gap between data workers and lay citizens in terms of their interaction and understanding of raw data since raw data is not easily understood by all people [16]. Participants also suggested that **visuals** or infographics are also good alternatives for data presentation. In ODPH, users can display data in a graphical form but only one participant was able to discover this and it took longer than expected to load.

DISCUSSION

Our interview findings corroborate the initial findings we had from our survey. In terms of the information seeking behavior of citizens, regardless of data experience and level of personal involvement, participants from both studies still relied heavily on Google to find needed data or information as seen in previous studies [10]. Although some participants believed that some of the available data from the portals are useful and relevant, most of the available data were mostly incomplete and outdated. This could be mainly attributed to the pace of digital transformation of the Philippines and the capability of the government to adapt to new technology to better manage and publish data. To address this issue, strict policies on data management and awareness campaigns are needed. We believe that design improvements could also be made to the current open data portals as majority of our participants shared common design and interactivity concerns with both portals. Both our results and previous literature [8] suggest that existing search techniques from web searches do not apply to data search. Although some of the participants were able to adapt to the search engine, we believe that an **improved search engine** would enable faster and more efficient searches. As spatial and temporal keywords were common in search queries, we recommend that the search engine looks through the metadata of uploaded data. However, this would require proper documentation of published data which may only be controlled by the data provider. Our results also show that regardless of data experience, citizens prefer a straightforward answer when seeking information. To support this, we emphasize the need for **descriptive visualizations** to communicate the content of the data. The majority of our lay participants expressed interest in visual aids to help them understand what the values in the data mean. Although storytelling and visualizations have been determined effective in communicating data [6], further research is needed to understand how Filipinos perceive data visualizations. Citizens need to be able to know how to read graphs and visuals presented. As previous literature suggests, there are various factors that affect the attitude and perception of individuals about visualizations [12].

LIMITATIONS

We acknowledge that a longer-term study that examines the information needs and behavior of more participants could provide a more comprehensive understanding of the information seeking and sensemaking behavior of lay citizens. However, despite the small sample size, we were able to identify design elements that would benefit the information seeking and sensemaking experience of citizens with OGD portals.

CONCLUSION AND FUTURE WORK

We conducted a needfinding survey of 119 respondents to understand the information needs and awareness of Filipinos regarding the open government data portals provided by the

Philippine government. We also investigated the information seeking behavior of 13 lay citizens and 8 data workers to compare the behavior of both groups. In addition, we also conducted usability tests and semi-structured interviews with these 21 participants to identify gaps in the design of the two portals. We found that only some information needs of lay citizens are answerable by currently available open government data. However, these datasets only provide partial answers to the specific questions relevant to the citizens. We also found that lay citizens would benefit from localized and advanced search engines, visualized, contextualized, and processed content for better sensemaking. Our findings and implications emphasize the importance of a citizen-driven design of open government data portals to increase citizen engagement in open government initiatives. Moving forward, we would like to develop and test prototypes using the proposed design guidelines. We plan to conduct an iterative experiment with Filipino citizens to measure user engagement and also factoring the varying levels of engagement with information visualization as defined in [11].

REFERENCES

[1] Afua Bruce. 2019. Open data promotes citizen engagement at the local level. (May 2019). https://thehill.com/opinion/technology/439467-open-data-promotes-citizen-engagement-at-the-local-level

[2] Michael Canares, Dave Marcial, and Marijoe Narca. 2015. Enhancing Citizen Engagement with Open Government Data. *Open Data Research Symposium* (2015). http://www.opendataresearch.org/dl/symposium2015/odrs2015-paper15.pdf

[3] Joohee Choi and Yla Tausczik. 2017. Characteristics of Collaboration in the Emerging Practice of Open Data Analysis. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, New York, NY, USA, 835–846. DOI: http://dx.doi.org/10.1145/2998181.2998265

[4] Ian Nicolas P. Cigaral. 2016. EO on freedom of information takes effect today: Palace. (Nov. 2016). http://www.bworldonline.com/content.php?section=Nation&title=eo-on-freedom-of-information-takes-effect-today-palace&id=136876

[5] Sharon S. Dawes, Lyudmila Vidiasova, and Olga Parkhimovich. 2016. Planning and designing open government data programs: An ecosystem approach. *Government Information Quarterly* 33, 1 (2016), 15–27. DOI:http://dx.doi.org/https://doi.org/10.1016/j.giq.2016.01.003

[6] Sheena Erete, Emily Ryou, Geoff Smith, Khristina Marie Fassett, and Sarah Duda. 2016. Storytelling with Data: Examining the Use of Data by Non-Profit Organizations. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. ACM, New York, NY, USA, 1273–1283. DOI: http://dx.doi.org/10.1145/2818048.2820068

[7] Julien Hivon and Ryad Titah. 2017. Conceptualizing citizen participation in open data use at the city level. *Transforming Government: People, Process and Policy* 11, 1 (March 2017), 99–118. DOI: http://dx.doi.org/10.1108/tg-12-2015-0053

[8] Emilia Kacprzak, Laura Koesten, Luis-Daniel Ibáñez, Tom Blount, Jeni Tennison, and Elena Simperl. 2019. Characterising dataset search - An analysis of search logs and data requests. *Journal of Web Semantics* 55 (2019), 37–55.

[9] Laura Koesten, Emilia Kacprzak, Jeni Tennison, and Elena Simperl. 2019. Collaborative Practices with Structured Data: Do Tools Support What Users Need?. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 100, 14 pages. DOI: http://dx.doi.org/10.1145/3290605.3300330

[10] Laura M. Koesten, Emilia Kacprzak, Jenifer F. A. Tennison, and Elena Simperl. 2017. The Trials and Tribulations of Working with Structured Data: -a Study on Information Seeking Behaviour. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 1277–1289. DOI: http://dx.doi.org/10.1145/3025453.3025838

[11] Narges Mahyar, Sung-Hee Kim, and Bum Chul Kwon. 2015. Towards a Taxonomy for Evaluating User Engagement in Information Visualization.

[12] Evan M. Peck, Sofia E. Ayuso, and Omar El-Etr. 2019. Data is Personal: Attitudes and Perceptions of Data Visualization in Rural Pennsylvania. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 244, 12 pages. DOI: http://dx.doi.org/10.1145/3290605.3300474

[13] Aare Puussaar, Ian G. Johnson, Kyle Montague, Philip James, and Peter Wright. 2018. Making Open Data Work for Civic Advocacy. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 143 (Nov. 2018), 20 pages. DOI:http://dx.doi.org/10.1145/3274412

[14] Francois Van Schalkwyk, Michael Canares, Sumandro Chattapadhyay, and Alexander Andrason. 2015. Open Data Intermediaries in Developing Countries. (6 2015). DOI:http://dx.doi.org/10.6084/m9.figshare.1449222.v1

[15] Mara Warwick. 2017. Philippines: Open Data Launch. (March 2017). https://www.worldbank.org/en/news/speech/2017/03/02/open-data-launch

[16] Vishanth Weerakkody, Zahir Irani, Kawal Kapoor, Uthayasankar Sivarajah, and Yogesh K. Dwivedi. 2017. Open Data and Its Usability: An Empirical View from the Citizen's Perspective. *Information Systems Frontiers* 19, 2 (April 2017), 285–300. DOI: http://dx.doi.org/10.1007/s10796-016-9679-1

[17] Harlan Yu and David Robinson. 2012. The New Ambiguity of 'Open Government'. *UCLA law review discourse* 59 (2 2012). DOI: http://dx.doi.org/10.2139/ssrn.2012489