TXTDS503

Paul Jason Perez

Dataset Biography

* Who collected (or will collect) the data?
  + I am the one who will be collecting the data.
* Where can the data be found? (Please provide a link if available)
  + The data can be found at the Philippine eFOI Portal at <https://foi.gov.ph>. Specifically, I will be looking at FOI requests that contain the keyword “SALN” which stands for Statement of Assets, Liabilities, and Net Worth.
* How was the data collected? What significant interpretations or choices went into the collection and categorization of the data?
  + The data will be collected through web scraping. When doing web scraping, I am mindful of the traffic that I might cause to the server, especially that this is a government service and there are people who need to use it. I am guided by ethical web scraping from Densmore (2017) which list some practical considerations when doing web scraping. I have written a Python script for web scraping the eFOI portal, which is available at my GitHub repository: <https://github.com/pvperez1/efoi_webscrape_script>.
* What is the social and historical context of the data?
  + SALNs are standard government forms in the Philippines that is required to be submitted by every government employee and elected officials annually. The document is supposed to be publicly available as a way for the citizens to check that government employees and elected officials do not live a lavish lifestyle that is beyond their means. Ideally, it is a document to check if an elected official is gaining wealth from the public coffers. However, in practice, requesting for such document is a highly political and complicated process in the Philippines. For example, the Philippine Center for Investigative Journalism (PCIJ) has requested numerous times for the SALN of former President Duterte, only to be denied again and again (Teodoro, 2021).
  + Thus, having a dataset on the FOI requests for SALNs can lead to further analysis on the process of obtaining the SALN
* What background information or domain knowledge is necessary to use this data responsibly?
  + An understanding of what the SALN is and what it is used for. An FAQ was written by the Philippine Department of Science and Technology that is [available online](https://asti.dost.gov.ph/wp-content/uploads/2019/11/ASTI-SALN-REVIEW-AND-COMPLIANCE-PROCEDURE-and-FAQs.pdf).
  + It is also important to understand that requesting for SALNs is a highly political and complex activity. This short report by a Philippine journalist is an easy read that can give users an idea of the process: [Why is it so difficult to access SALNs of House lawmakers?](https://www.rappler.com/newsbreak/iq/210092-why-difficult-access-salns-house-lawmakers/)
* What are limitations or problems with the data?
  + There may be some requests that were written in Tagalog (or one of the many local languages in the Philippines). Thus, familiarity with the local context will be advantageous.
* What ethical considerations must be kept in mind when using this data?
  + When using this data, it is important to focus on the requests and not the requester. Below is a snippet of what can be seen and extracted from the eFOI Portal:

A screenshot of a chat

Description automatically generated

It can be noted here that we can see the name of the requester, even though abbreviated, which in this example is J. Ribaya. Furthermore, these are the legal names of the citizens requesting for such document because they need to provide a government ID when signing up for the eFOI portal. Thus, if we know that a certain reporter has published an article about a certain politician’s SALN, and we are familiar with their initial and last name, we can now monitor what FOI requests they are submitting.

* What other information is essential to understanding and working with this data?
  + Aside from an understanding of the basic mechanisms of Freedom of Information requests, it will be helpful for user to browse the eFOI portal and observe how citizens usually file their requests. For example, it can be noted that for some people, they use the eFOI portal to request for government services or customer support (which are both deemed as invalid FOI requests) (Perez, 2020).

References:

Densmore, James. (2017). Ethics in Web Scraping. *Towards Data Science*. Retrieved from <https://towardsdatascience.com/ethics-in-web-scraping-b96b18136f01>

Teodoro, Martha. (2021). PCIJ request for Duterte SALN runs into dead end – again. *Philippine Center for Investigative Journalism*. Retrieved from <https://pcij.org/2021/07/26/duterte-saln-restricted-pcij-request/>

Perez, P. J. (2020). Why are Freedom of Information Requests in the Philippines Being Denied? An Analysis using a Practice-based Approach. Philippine Journal of Librarianship and Information Studies, 40(2), 3–36. Retrieved from <https://phjlis.org/index.php/phjlis/article/view/41>