

My initial plan was to examine my YouTube data and though this data, see if it is possible to derive a situated memory or meaning from it. Although Google is offering a service called Google takeout, where the user can download a lot of different categories of data and amongst those YouTube data, the format the data is outputted is not very suitable for examination. When you request your personal data from Google, depending on the size of the dataset, you will receive an email containing a link to the data. After extracting the data, I realized it is comprised of different HTML files, such as comments, watch history, search history and subscriptions. This makes the data well represented visually for personal use, but ill represented for analysis. Before anyone can derive meaning from this data, it would have to be comprised of more than just the search query and the names of the videos the user has watched. If I had access to the metadata ascribed to the videos, I would be able to categorize them into themes and with the date of watching assign a mood/situation to that specific date. I am uncertain about the reasoning for Google to make the extracted data maladroito to work with. In some ways it makes sense for them not to assign their own and other users metadata to the extracted data. These metadata are not something I as a user have had anything to do with and therefore have no right to access. With all this said I did end up using Googles takeout service for my data-selfie. Because I own a Google phone and have tracking activated on my Google maps application, they have access to my phones every move for the last three years. This data is also available from the takeout page and came in a more suitable format for a data working environment. The data is divided into folders with years and within those folders there is JSON files for each month. I had in mind to visualize my location history on a map, possibly a heatmap showing my most popular locations. This would help proving my assumption, that I almost never leave my apartment. This location data can also be used in tracking and preventing stress and depression. As Alex Pentland proves in his paper “Human Dynamics and Mental Health” you can prove with a very high accuracy, the risk of depression from only location data. This is due to the fact, that for a person to maintain a good mental health they must have regular interactions with other people. Karl Marx said something along the line with “*society is the sum of all our social relationships*”. And Adam Smith said: “*it is human nature to exchange not only goods but also ideas and favors (...) it is these exchanges that create solutions for the community*”.¹ With this location data of people in an area you can see how often they interact and from this predict depressive tendencies. Back to the data I have in my JSON files and what I can actually do with it. The file can usually be parsed into a CSV file, which then can be made into a map using carto.com, an application Pablo showed us. In the case of the Google takeout data, this is not possible. Again, I am unsure about the reason, but I assume it is because there are too many categories in the file. It takes parameters like endLocation, startLocation, duration, distance, latE7, longE7, activityType, placeVisit, placeConfidence, timestampsMS and

¹ Alex Pentland, Human Interaction, Idea Flows, and Wealth Generation, 2018

editConfirmationStatus. For my visualization I only needed the coordinates and the timestamps but extracting just those datapoints got tricky, because the datapoints are very interwoven. Using the interactive data visualization software Tableau I managed to isolate a category called simplifiedRawPath, that contained the parameters latitude, longitude and timestamps. I was not able to combine all the files; that is why the maps are divided in months like the files I extracted was. To represent the different times, I made all the timestamps different colors. This makes it possible to calculate the timestamps into a time and date humans can read, so for example the timestamp 1597762482 is converted to 18/08/2020 at 16:54:42 and I can find that location on my map and see that I was in the botanical garden at this time. From working with the data, I encountered a lot of platform-standardization issues, where the data I wanted to work with either did not come in a format I could work with, or it was too interwoven like the location data. Nanna Thylstrup talks about the infrastructures of mass digitization and how standardizing processes are at the core of this.

“The infrastructures of mass digitization are thus media of politics and politics, at times visible and at others barely legible or felt, and home both to dissent as well as to standardizing measures”²

She goes on calling standardization a mode of capturing, conceptualizing and configuring reality. I am therefore left wondering if it is my lack of expertise in converting files into suitable data sources or if the provider of the data (Google) is actively making it harder for other actors to analyze the data, keeping themselves a step ahead of all competition. This is at least true for companies like Twitter, where you can only query a limited number of tweets (1% of all tweets). Initially I would think that platforms do this to maintain a competitive advantage, but Amelia Acker and Adam Kreisberg argue against this by saying:

“Platforms such as Facebook, Twitter, and Pinterest do not make their data interoperable because their competitive advantages over each other lie in their differences, not their similarities. Actions such as posts, likes, replies, reposts, and comments each have their own definition on each platform and are reflected differently in the offline representations of content creators’ data”.

You could argue Michel Foucaults notion of normalizing power in context to me providing Google with my whereabouts at any time. They do not force me to have location shared, they do not punish their users for denying them location data. They willingness to share the location inhabit the user, because it enables more functionality from the device. So if we take an everyday example of how I have personal use of

² Thylstrup, Nanna Bonde. 2018, The Politics of Mass Digitization

visualizing my location data, it could be the case of forgetting where a bike was parked. The map allows me to travel back in time to the last time I was riding a bike and then locate the end of this bike ride. Personally, I would have lost my bike more than a few times if it was not for this functionality. The question comes down to the value of this location data and the use cases that hopefully differ a lot from me to Google.