Project One:

Data Detox

IT University of Copenhagen

Autumn 2017

Written by:

Israe Noureldin
Islam Jabrayil Mastanov
Richard Banyi
Soltan Reza Hoseini
Tom Roberts

Introduction

In the contemporary society where the internet has transformed the foundation of business and economics, telecommunication, transportation and basically every other major industry, data becomes one of the products of such transformations, therefore the technologies around Big Data are a vital element of this network infrastructure.

For that reason, there has been a massive attention towards the protocols in which the Internet is facilitated by, and towards the different technologies around data storages, data mining processes, data analytics, developing algorithms and machine learning, social media and so on. As we enthusiastically move into the future and embrace different digital tools and technologies such as smartphones, smart TVs and Chromebooks and Ultrabooks, we learn that there is little notice among general population regarding data privacy and data protection, especially when one might learn about the amount of data being produced from people's activities across different social media channels and platforms such as Facebook, Instagram, Snapchat etc.

This paper is a brief report and a group-reflection upon practices around awareness of data privacy, data ownership and power struggle in the age of the Internet. A report about individuals' rights to control their own data and accept social responsibility for their data practices.

Discussion

Before starting this section, we believe it is important to mention that instead of using our notes from Data Detox week, we decided to use our notes from the Data Detox discussion. We spent half a day to discuss and compare our experiences throughout the Data Detox week and that has lead us to some interesting points and questions.

With the deep implementation of the Internet in the foundational layers of society and economics, it is almost impossible to imagine a world without internet especially when this technology has become so efficient in improving our lives in many ways such as knowledge and information sharing, health care, education, environment etc.

During our Data Detox discussions, we have come up with two major approaches when we discussed power struggle in the age of Internet and data ownership, data accessibility and accountability.

Two approaches:

A. Some giant companies such as Facebook, Amazon, and Google dominate the market advertising and cloud data storage and that results in some corporations becoming very powerful, which raises concerns about the future of the Internet [quote Soltan].

B. In the contemporary society, in every field and industry, there are a number of powerful players, and if Google get so powerful, it doesn't affect me as long as I know they are not using their power in a wrong way [quote Israe].

Both approaches deserve further discussions, because both open new approaches and concerns towards the future of the Internet. One simple and obvious question that came up was why these companies are collecting so much information. Well it's deeply rooted in their business model.

As Shoshana Zuboff popularizes we live in an age of Surveillance Capitalism, she states that "the online world, which used to be kind of our world, is now where capitalism is developing in new ways". Moreover Nick Couldry believes that "surveillance capitalism is focused on data extraction rather than the production of new goods, thus generating intense concentrations of power over extraction and threatening core values such as freedom and privacy".

The primary goal of social platforms, advertisement industry and all other digital products/services is to keep capture our attention, to persuade us to behave certain way, to buy one more pair of shoes in the context of profit making. By capturing our usage of their services, it allows them to survey and shape opinions with machine learning that scale. Computation can figure our identities and moods. With all this power it is a perfect set up to totalitarianism. But I would argue that Silicon Valley companies are still kind of liberal if we take on other hand China. The key observation is that all the persuasion is happening in the dark, behind the scenes, and it is individually targeted not as a group on open space. On TV companies advertise to everyone, but on the internet it's fragmented and profiled. We don't see what other people are seeing. Which gives the power to companies to figure out what people are up to.

This might present a danger. Certain organizations might take advantage of it and misuse it in their own selfish-intensions. An example for this could have seen in the recently American elections when the Russians misused Google/Facebook algorithms and produced fake news and managed to manipulate/pursue the voters. Therefore we think the key point is the realization of the manipulative intentions of these services.

However, we think persuasion is just persuasion. It may not be a problem. There are times when we're fine what captures our attention. We have to understand the dangers that algorithm can impose and have to make our own decisions Whatever power we grant these companies we have to understand what will be inherited by them and how and when companies should have access to our private information.

Another commonality that came up was about the 'right to be forgotten'. Several of us had produced data on Facebook and Instagram in the past that at the time we felt was important. Having matured and having changed opinions since then, it was no longer important to us, and we wanted to remove it. It was possible to go through posts, photos etc. one-by-one and remove them, but this was tedious and time consuming. We used a chrome extension called data-selfie which gave us a really good overview what Facebook knows about us. Unsurprisingly the Facebook algorithms were able to gain lot of insights about us, for example religious and political orientation, shopping, health and other preferences. What surprised us the most is the fact that Facebook even store data that we did not published. We were able to see the data that we have inputted in the post or comment field without submitting [quote Richard]. Similarly, if we wanted to change or delete a social media profile, we found that search engines like Google still had cached versions of these old profiles in their results that were being shown [quote Tom]. This led us to reflect on how the data that is important to us now may not be important to us in 10 years' time. This suggests that the set of data we consider important is not fixed and is in fact always changing, and the right to edit or delete data is very important. Furthermore, while we can delete and edit our data, this may not impact any analyses that companies have already performed on our data. This way, we may be able to edit our 'user-facing' data but we have limited access to our 'inferred' data.

Another topic we have discussed was deleted data; when one's data is deleted, is it deleted from the internet? If one deletes a picture or a tweet from their accounts, is it deleted from the databases? How can the deleted data affect the profiling issues we are facing? Digital services we use today are not designed by default to give option to users to delete their data, and it is not given the permission to users to completely wipe out their history of digital interaction.

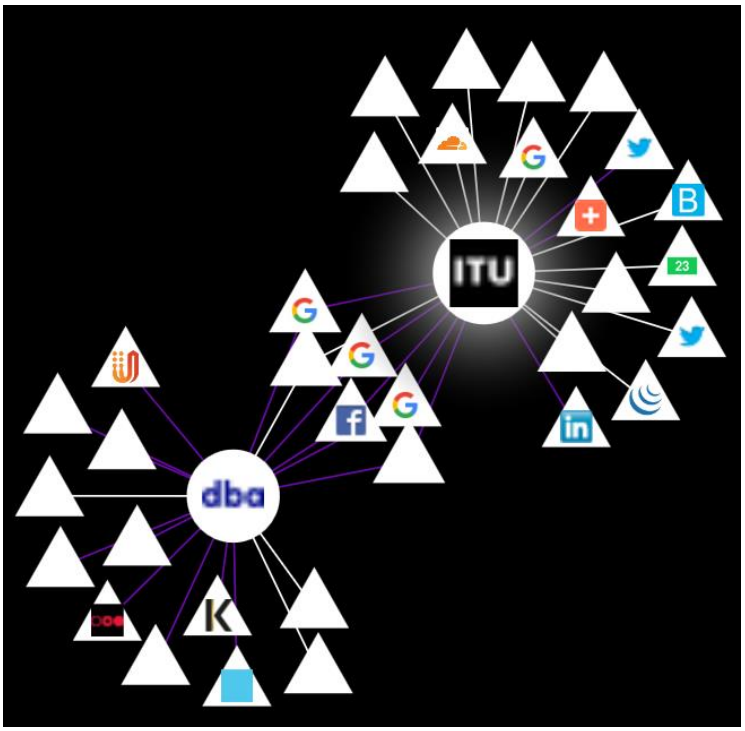
This can hint to the opinion of Richard Stallman on using digital services, as he puts it: "If you use a proprietary program or somebody else's web server, you're defenseless. You're putty in the hands of whoever developed that software."

Interestingly multiple members of our group created and start using their Facebook once enrolling in universities as it was means of communication between other students. This is another example how it is almost impossible to participate socially in society where these platforms became a status quo. Although

members had different reasons for not using Facebook prior to starting the university, one of our member specifically avoided its use due to his concern over Facebook's proprietary use of data. Going through the Data Detox and being familiar with most of the tools and techniques provided it was surprising for me to see the limited use of them among my constituents who were not aware of the privacy violations they are subjected to [quote Islam]. Elaborating on this with our group has led us to presume how large can be the portion of world's population that are completely unaware about the data being collected on them.

Data-fasting and Opting-out are not a solution

With the internet rooted so deep in the developed countries' infrastructure, it is impossible to stop using digital services as Janet Vertesi experienced that trying to keep her pregnancy condition as a secret made her confront her family members and friends and even a suspicious member of the society. Google and Facebook dominate the digital market and by going online and use digital services, in one way or another we interact with these companies especially Google, as the Lightbeam showed us in the day 6 of Data Detox, that how Google and Facebook know where you are and how you are interacting with digital services, even when the user is not interacting directly with these services.



As we also experienced during data-fasting session, it will become a pain if we stop producing data. So this hints that it is very important to seek for the solutions to protect personal data, especially in an era where there is more focus on advancing technology than discussing what privacy is and how to protect one's privacy.

Databox

We understand that the goals of the Databox are interesting as it tries to find a compromise between the 'all or nothing' situation where users either forfeit their privacy, or forfeit their use of services that have become deeply ingrained in modern society. It aims to give attractive incentives to both the companies wishing to use people's data, and the data subjects themselves to find a middle-ground in data privacy. Databox's ideas about giving alternative incentives to both third party companies and users is particularly interesting as this gives the option to have this middle ground: those who want to continue to use services as they are today can continue to 'pay' using their privacy, whilst others can preserve their privacy and pay for services using money. This could allow privacy to be preserved while remaining engaged with modern, networked services.

It's clear that there are still some major technical hurdles to overcome, and previous attempts have had differing levels of success. Storing all of a person's data in one place makes it essential that it is secure and always available. Managing this single point of failure would likely be a major problem to resolve. Even if the infrastructure is in place, it is not clear how early adopters could be encouraged to get onboard with the system as a service like Databox would only truly work according to the network effect whereby it is only attractive if it already has parties using it.

We also think that giving this power directly to users could be beneficial psychologically as subjects feel empowered that they have fine-grained control over what is known about them, and how this knowledge is used. This could shift the focus from consenting to hand data over, to giving consent to having certain conclusions drawn about you. However, to even reach this point, first collecting and merging data from the plethora of existing (and often proprietary) sources is a difficult task, and keeping this information updated and synchronised across devices adds to the complexity.

It is very interesting that an important point of Databox is usability. It's almost useless to give this level of control to subjects if they are not able to understand exactly what their data says about them, and what third party services can infer about them. However, we questioned whether the average user has the capabilities and knowledge to act properly when given this power, for example, knowing how to recover after a data

breach. Moreover, we discussed how having the power to edit and delete the entirety of our data could possibly exacerbate the 'perfect record' issue often debated in social media settings.

However the concept of the Databox is still vague for us regarding the interaction of users on giant digital platforms such as Google and Facebook while owning a Databox. The main question one might ask is that how Databox can help to protect one's data against the giant data corporations. For that reason, we don't assume that Databox can make a user in charge of their data that are owned by Twitter, Facebook, Google or Apple.

In the paper "A critical review of 10 years of Privacy Technology", Danezis and Gurses describe privacy as confidentiality: "privacy is hence defined as avoiding making personal information accessible to a greater public. If the personal data becomes public, privacy is lost. Moreover privacy is described as control: "the right of the individual to decide what information about himself should be communicated to others and under what circumstances".

If we apply the lens of privacy as confidentiality to Databox, it technically provides privacy to individuals because users decide whom to share their data with and for how long, it keeps users in control of their data even when they granted access to a company, they still have the option to stop this access.

If we look at the technology through the second lens which is privacy as control, this is still a good solution to choose what information should be communicated to whom and under which circumstances. However, as we discussed earlier, we still haven't understood how Databox can challenge the fact that giant companies have access and store users' data. Short said, this technology can provide a ground to have control over your data against smaller companies, not so much against big companies where we pay for the services by our data, so in that sense it can only be a partial solution.

An important concept of the Databox seems to be to provide a way to decentralise how data is stored. If 'data silo' companies continue to store private data and decide how it is accessed and used, any conflicts of interest these companies have with third parties could have implications on the users whose data they are storing. Providing a decentralised option could take away this control and limit negative effects in this regard.

Conclusion

Technologies such as Databox, regulations such as GDPR, and practices such as opting out of the digital world are not ultimate solutions, but these partial solutions are needed to engage the general population, to encourage innovators to focus on this issue and build upon them. They are essential to move forward, to

appropriate further innovation and lay the foundation of a future where people are more aware and capable of protecting their privacy and personal data.

We need to move towards a future where digital services by design are meant to give privacy and protection to individuals, a future where the answer to privacy is not Databox or opting out.

It is true that each individual needs to accept and be aware of their digital behaviour and be accountable for what they do, and be aware that what are the consequences of paying for digital services they use by their data, but on the other hand, service providers by design limit the amount of control each user has over their data, so the questions remains to be: how much control over our digital presence we acquire even if we accept the responsibilities that come with using digital service? Should one stop using the Internet services completely? What happens when individuals are forced to use digital service to interact with government to for example to fix their taxes or to use the social and health services? Where is the line between accepting responsibility and taking control drawn?

To summarize, we conclude by emphasizing that this topic is so broad and at the same time can be so cloudy and vague that takes time and effort to solidify these concepts and draw the lines between different concepts and definitions such as privacy, data protection or taking control of personal data. This topic needs more attention, needs more discussion and innovation, and hopefully with these engagements, we can see a better and more private future in the era of Internet for each individual.

References

Couldry, N. (2014). The price of connection: 'surveillance capitalism'. Assessed at: https://theconversation.com/the-price-of-connection-surveillance-capitalism-64124

Danezis, G., & Gürses, S. (2010). A critical review of 10 years of privacy technology. Proceedings of surveillance cultures: a global surveillance society, 1-16

Stallman, R. (2008). Cloud computing is a trap, warns GNU founder Richard Stallman. Assessed at: https://www.theguardian.com/technology/2008/sep/29/cloud.computing.richard.stallman

Vertesi, J. (2014). My Experiment Opting Out of Big Data Made Me Look Like a Criminal. Assessed at: http://time.com/83200/privacy-internet-big-data-opt-out/

Zuboff, S. (2014). Reality is the Next Big Thing: Keynote. Assessed at: http://davidcharles.info/2015/01/shoshana-zuboff-surveillance-capitalism/