

# Portfolio: Curating Data

Student: Jeppe Mark Vind

Student number: 201805354

University: Aarhus University

Course: Curating Data

Hand in: 2020-12-21 14:00

Characters incl. spaces: 34.209

## Contents

Contents .....	2
Introduction.....	3
Assignment 1 .....	4
Assignment .....	4
Assignment 2 .....	7
Introduction.....	7
Creating the data and graphs .....	7
Analyze .....	9
Implications .....	9
Ethical considerations.....	10
Assignment 3 .....	11
Introduction.....	11
Methodological Considerations.....	11
Hierarchy .....	11
Zotero and its Software Limitations and Possibilities .....	12
Tags.....	12
Assignment 4 .....	15
Level 1.....	15
Store .....	15
Plants.....	15
Humans.....	15
Time .....	16
Level 2.....	16
Advertisements.....	16
Cellphone.....	16
Surveillance.....	17
Bicycle Rental.....	17
Curatorial Statement .....	18
Conclusion .....	19
Appendix.....	19
Bibliography.....	20
Icons from the webpage.....	20

## Introduction

Through this portfolio, I will go through 4 different assignments made in the course “Curating Data” on Aarhus University in the fall of 2020. Assignment 1, 2, and 4 has been made by me solely, while assignment 3 was a group assignment made with Mads Vestergaard and Sia Kofoed-Heller.

In the portfolio itself there is no appendix, but it is accessible through my website, which is also used for assignment 4, at <https://vindvind.eu/curatingdata/>. I highly recommend that you use Google Chrome to access the website. Before reading the segment “Assignment 4”, I am encouraging you to visit the webpage, since the experience might be altered by not doing so. The segment “Assignment 4” consists of the various line of thoughts I have had throughout the course, which ties the assignment and the whole course together for me.

The overarching theme for me, has been how objects change their representation, either through digitization, classification or situation. Assignment 1 makes this clear, as a physical element becomes digital, and to me, becomes something else. In the physical state the only traits I desire are how they look, and whether they are in good health. This is made difficult through digitization, as it loses these traits. Health wise since it is disconnected from its physical counterpart, and visualization since it is stored within a database. You could seek to combine them, but that would rather be a design focused task, rather than this assignment, though it shows an interconnection between the two.

This line of thought can be continued in assignment 2, where I was wondering what the “data selfie” I perceived could be interpreted as. I could see the changes in my own life within the messages, but I was left wondering what it would be through Facebook’s point of view. The last question can however not be answered due to Facebook’s opacity, or lack thereof. This is a common theme in Facebook, which also include me being categorized into “interests”, without me knowing. Taking assignment 4 into consideration, this proves a major issue, due to the difference in ways of perceiving a tag, meaning there is a high risk of being mislabeled. All these factors make up for my “dividual” (Cheney-Lippold, 2017), which influence me knowingly or not.

I will further expand on this concern in assignment 4, of who I am digitally, and how this influence my actions, my experience, and myself entirely.

## Assignment 1

### Assignment

#### *Introduction*

For this assignment I have chosen to digitize the plants in my household, for several reasons. First of all, in order to keep track of when my plants have been watered, as to not water them too much, and secondly, to know the names of each plants.

#### *Creating the data set*

In order to digitize the plants, I first took a photography of each separate plant, using my cellphone, which then through the application "PlantNet", I managed to determine each plants specie, family, genus and common name. Secondly, I put in additional information, in order to create a "usecase" for this dataset. In this instance it was managing water, age, and a picture of the plant.

#### *Implications*

This was a brief introduction as to why and how I created the attached dataset, and through this introduction, you can immediately tell, that the data here is not "raw". As Gitelman & Jackson elaborate on in "Raw Data Is an Oxymoron", there is no such thing as raw data, though it may be implied by medias. This data is highly biased, not only in the information shown, but also how I have chosen to display the data. For instance, many of the plants have more common names than shown in the dataset (according to the application PlantNet), but I have chosen not to display them, since I either though they were highly regional names, or that they were duplicates of other names. This also leans into Haraway's text of "Situated Knowledges", where it becomes obvious, that each text (in the phenomenological meaning as well) is situated, meaning that it is subjective. This is relevant in the dataset both in the gathering of the data, where my (and the crowdsourced names of the plants through PlantNet), is subjective in its own way.

Even the machine learning algorithm, which found the names of the plant, is subjective in its own way, since it is fed data from the users. The whole "feeding process" of the algorithm is as well very biased, in the sense that you rely on other users, or developers/researchers, to categorize the plants correctly (from your perspective), in order to get the desired result. This perspective can be translated into "power" in more senses. First of all, this empowers me to be able to actually retrieve the names of my plants, which eases the caring of the plants. Secondly, this algorithm, and all the other users indirectly, holds power over me, influence of sort, in the form, that they are able to impact the suggestions of plants I get when searching.

The whole idea of power in digitization, is also something Thylstrup touches in "The Politics of Mass

Digitization" (Thylstrup, 2018), where it become obvious, that being able to influence the output of algorithms are quite powerful. Not only that, as many texts elaborates on, just the sheer fact of understanding algorithms and computation is a power in and of itself, which separates this very generation!

#### *Ethical considerations and perspectives*

As stated earlier, I have chosen some of the common names to show in the data set, and hence also chosen not to include some. This was not due to them being unethical. Actually, it was more the other way around, where I have chosen to include some which could be interpreted as unethical. This for instance being "The Wandering Jew", which to some might be offensive. This consideration I chose however not to act upon, since no matter how offensive the name might be, it is a common name. This relates to me stating in the beginning, that I chose to create this dataset for myself, hence it is not a public record. I chose to include these, since it may help me explaining others which plant I am talking about, and search more broadly for the plant online. With this argumentation, I am very well aware, that I too constitute the name, by taking it in, and using it. Despite this, it is easier to search when having a broad variety of names, and also this task was to explore what could be included, and the ethical considerations here of, which also is a reason, that I have chosen to include this very name.

This is not me saying, that personal use trumps ethical consideration, but more so a statement that it is a delicate balance between ethical consideration and use of the database. This is especially true taking all ethical considerations into play, which would include all individuals in the known world. I am therefore inclined to take Immanuel Kant's stance on ethics here, by claiming that this action is not unethical in itself, since I actually would want anybody in my shoes to make the exact same actions. This of course not meaning that is in an ethical action, since I am not doing it in and of the ethics itself.

There is also a question to be asked, both through this assignment, but also within the task: Which things should we digitize, and what implications might this have for the future?

My example, with digitizing plants, might for example be a step into a "Smart Home", where most things are controlled by computational devices, that I may have some degree of control over. This leads to the phenomenon of "Techo-Chauvanism", which was especially on its rise in the late 90's with the "dot-com" era, where everything seemed to be digitized, and the question most often asked was not "should this be digitized?" but rather "why should it not?". As with all ethical considerations, it is possible to completely devastate the use of the object completely through ethical considerations, but I believe we own it, both to our self, but especially our future society, to ask whether it is really necessary to digitize everything. An argument could be made, that these physical objects, such as my plants, are digitized in order to preserve them, but the question that could be asked here is, whether the ones in the table are actually my plants, a

representation of them, or something completely else?

They become what you could describe as “Evocative Objects” (Turkle 2007), meaning they carry with them a new, or rather amended meaning. Hence they are my plants, but still they are not, since they have now become digital, and can shift ownership, and even be copied, without direct consequences to my plants. It is a fun thought experiment; how these plants would transform, what they would become, if they were to be copied, examined, and understood by someone else completely. Fun until you realize that it is what is happening with humans in the algorithms driven by surveillance economy (Zuboff 2015). What become of us, if we were to take the place of the plants?

## Assignment 2

### Introduction

Communication is increasingly going through the world wide web, instead of the traditional analogue forms. Who can for instance remember the last time, they picked up a pen, and wrote a letter to someone, instead of simply going to WhatsApp, Facebook Messenger, Outlook, or something completely different? In this assignment, I have chosen to focus on my messages across Facebook Messenger from when I first started using Facebook, June 2008, and until present day.

This I did in order to see, if there was any particular pattern, from which I could extract some knowledge of the platform, the use of it, or maybe about myself.

### Creating the data and graphs

In order to investigate my messages of Facebook, I first had to retrieve my Facebook Messenger “history”. This is fortunately quite easy due to GDPR, where I could simply download all my messages ever sent from Facebook.

After receiving this 2.5 GB large .zip-file, I unpacked it, and began looking at the structures. From here I could create a script in Python, which went through all the messages, filtered them, and turned them into a CSV-file with only two parameters; date and amount of messages sent, which can be found under appendix 2.4.

Python is great for these types of task, since it is a light-weight program, which rapidly can go through all the messages. Furthermore, it is great at accessing JSON files, which your Facebook data is, if you do not choose HTML format.

After loading the data into the Python script (appendix 2.3), I looped through all the messages, and cut away all messages that was not from me. Secondly, I filtered away message that were automatically generated, such as “You can now contact X” and so on.

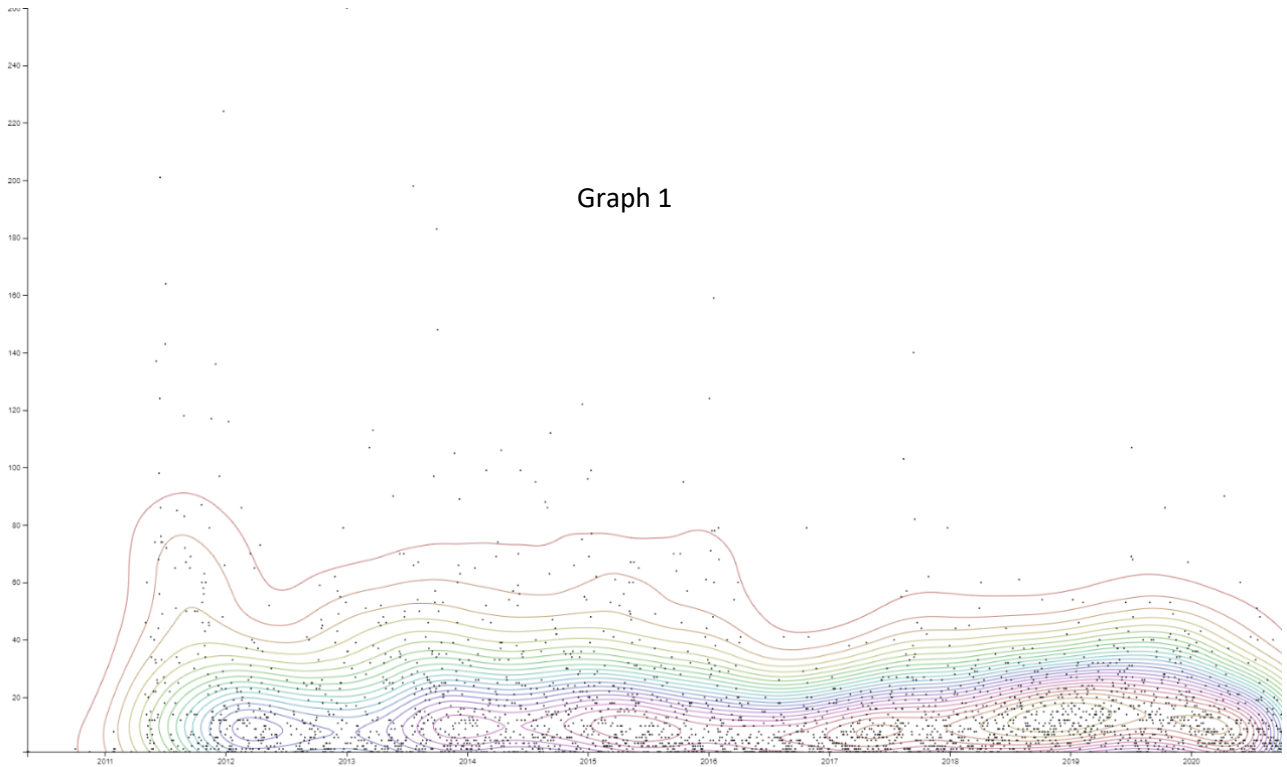
Afterwards I created a dictionary in Python, which is basically just an object with keys and values, in which I generated the data while looping through the messages.

After the dictionary had been completed, I wrote the data to a .txt file, written in the form “date, amount of messages sent”.

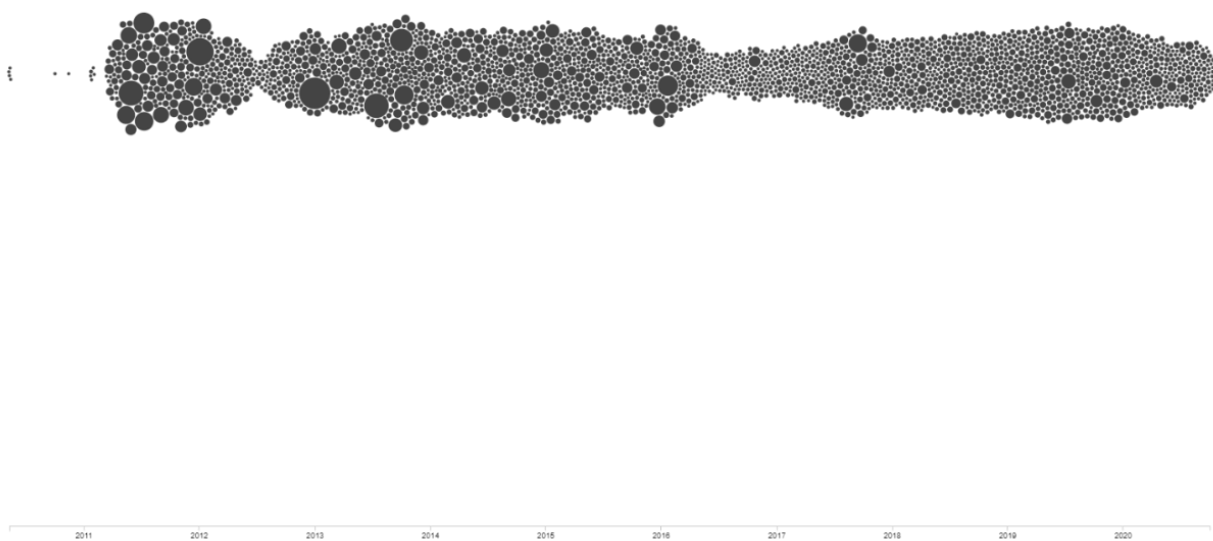
After this file had been generated, I was free to select a graph, which could display these data, so it was easier to perceive and analyze them.

Therefore I went to <https://app.rawgraphs.io/>, where you can generate graphs based on CSV files. I chose

the two found below and in appendix 2.5 and 2.6, since I found they both displayed the data easily readable, but in different ways.



Graph 2





## Analyze

The data shows, that I was most active on Facebook Messenger up until 2015, where I finished my high school diploma. The most messages were sent from 2011-2012, where I was on boarding school, which would indicate that I wrote a lot to the friends I had, who did not attend the same school. This is however contrary to the period of 2015-2018, where I was working abroad. This period was where I sent the least amount of messages, despite only being in Denmark approx. 5 weeks a year. This could be explained with, that in that period I did not have the need to organize gatherings in the weekend, or simply just different periods of life, where I was not so focused on being in contact with my network at all times.

It is however funny, to see how the amount of messages changes according to period of my life, where I can pinpoint events of life to the graph, and thinking about how it comes, that it changes the graphic as it does.

## Implications

This way of using data provided by Facebook, shows that the data is not only for companies to either exploit or sell, but can also be used to provide beautiful graphs, which serves to learn about yourself. The sheer way that this data is portrayed, as a company capital, is in itself an issue, as Kreisberg and Acker states. Why do we let these multibillion companies and their complex algorithms to undertake the task of not only controlling the data itself, but also its future? This has even more implications, as this data is part of the surveillance economy (Zuboff 2015), making me a dividual (Cheney-Lippold 2017). The critical questions are then: "How does these data represent me?" and "How can they be exploited?". This is at least the thought process through a critical lens. Through a more exploratory lens, you could ask, why I would have to go through all these steps, in order to visualize how messages, I send? Then we go right back to letting the companies process our data, sorting and labeling it for algorithmic rather than exploratory use, making it obvious, that the data is mainly fuel to their engine, and not for the users.

As Annette Markham would say: The biggest accomplishment by modern companies, to lead users to believe, that data is just a waste product, and that it should be "shared".

This point is also evident in Nanna Thylstrup's *The Politics of Mass Digitization* (2018), where it is pointed out how much power these companies hold.

An example of these powers, were also evident in class, where Pablo Gonzales showed a network of his Facebook friends, with data he had retrieved prior to 2016, where Facebook shut down the API, from which users could get data about themselves. It is, when you think about it, absurd, that users are prohibited access to their own data. And not only that, it is extremely difficult (and against Facebook Terms of Use), to retrieve them anyways (through automatization).

Why should it be this difficult and even against terms of conditions, to explore *our own data*?

### Ethical considerations

In this research I have excluded all names, and only used my own data, since I have not retrieved permission from all persons I have written to. This is crucial due to situatedness (Harraway 1988), since they are very well aware that the data are to be seen and used by me and Facebook, but have not willingly accepted the setting of which I portray the data in this paper.

Furthermore, I have only used data provided to me by Facebook, which have been given to me within the boundaries of Facebook's Terms of Use, since it is a grey area to tread, and putting my account in risk of being disabled. This just further underlines the power these companies hold, since I am not willing to put my account on the line, meaning Facebook has created a great "lock in" mechanism for its users.

## Assignment 3

### Introduction

The aim of this assignment was to build a collaborative collection of books and other published sources, which we have gathered in our group. We have chosen to use Zotero instead of Calibre, since we found it to be a more efficient software to work with collaboratively, and had earlier, good experiences with the software. We all have a part in both gathering the different texts and in writing the paper, since we sat together and discussed it.

We have imported our Zotero library and imported it to Excel for ease of reading. You can find it in appendix 3.2, and an image of the Zotero library in appendix 3.3.

### Methodological Considerations

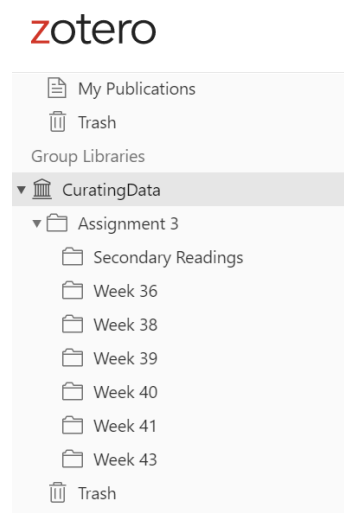
The texts chosen for our assignment are primarily literature from the course, supplemented with literature from our individual bachelor projects, which we are currently writing. These secondary texts of course have some meaning to us in the context of Curating Data, hence the reason we chose them. Furthermore, these texts are what we are currently working with, meaning there is a hidden gain in this assignment, in being forced, by this process, to think about our projects and texts in the light of Curating Data.

With this aim however, we ran into issues, since some of our texts were born physical. This does not mean, as investigated in assignment 1, that these objects cannot be transformed, but some value to these works ought to be lost in the process, though new ones may be gained.

### Hierarchy

As to categorization of the texts, we first decided on this hierarchy (see picture and appendix 1.1) since it made a clear distinction between self-chosen texts, and the ones provided by the course. Each folder contains the readings for that specific week provided by our teacher. The folder with Secondary Readings are texts that we have gathered from other courses, our individual bachelor projects and other texts, that we believe are relevant for the course.

This was done to easier navigate and sort the texts. This does not necessarily bring meaning to others than us, but as said in “Curating Online Exhibitions” (Connor, Michael 2020): “Digital culture is ‘more about practices than objects.’”, meaning that the hierarchy, and tags of course, has to be bound in our own understanding and practices, since we are the ones working with these texts.



## Zotero and its Software Limitations and Possibilities

Zotero as a software is used to share and cite texts. The only hierarchy present of Zotero would be its folder layout, meaning you can make some categories subordinate to others. We used this function, as shown in the previous section. The tags itself in Zotero are however not hierarchically divided. This makes for a great experience in our opinion, abeling us to divide the text both hierarchically and not. To further expand on Michael Connor's quote about digital objects, the non hierarchically part of Zotero makes it possible to search for tags, in this instance of the text's content, and making it possible to go back, in our instance to a specific week, and check what this week's courses was all about. However in the process of tagging and categorizing our texts Zotero affords us several built in features, whenever importing a text it will scan the metadata that is already connected to the PDF or EPUB file, which it will then try and import. Whenever this process is successful we end up with a file that has saved us some time on tagging, with the tags we've already decided to use, but further inspired us with tags that according to the author or distributor of the text, find useful. These tags are not always what we want to portray with our collection, but as previously mentioned, they can help inspire and improve on the quality of the tags we have chosen.

## Tags

We've chosen a wide variety of tags in order to categorize the texts to the best of our ability. We've already utilized the hierarchy and sorted the texts into different folders. However, in order to specify each text even further, we've come up with the followings tags:

Algorithms - A well defined sequence of specific instructions

Curating - a selection, presentation and organization of things

Social Media - Digital tool to share content and messages with other people

    Facebook - A social media website

Business - Commercial activity

    Marketing - Business action of promoting and selling products

Politics - The governance of countries or areas, also associated to power

Art - Various expressions of creative activity

Big Data - Collection, Storage and Analysis of large amounts of data

    Datafication - The process of more and more things being turned into data

    Dataveillance - Surveillance and collection of online data

    Metadata - Data about the data

Digitization - The process of making artifacts digital

Machine Learning - Computers that improve through experience

Cognitive - Relating to conscious intellectual activities, for instance thinking

Perception - The way in which, or the ability to, understand and interpret things

Feminism - A belief in women's rights based on the equality of the sexes

These tags were all chosen because of their relevance to the texts' content, and how they could potentially help us create a library of texts, where we could easily access information concerning any subject regarding the course. The process of coming up with the tags was a process that evolved as we worked with it. We had come up with an original list, which we then expanded upon as we were tagging each text, realizing that there were several other tags that the different texts could make use of.

The sheer tagging of the texts also made for some debates on whether one category was subordinate to the other, or if they held the same meaning. This only serves to further underline that tagging and curating in itself, is a biased and subjective task, which does not necessarily bring the same meaning to an outsider, as it would to the curator itself. This is the difficult part of creating any database, as tagging is essentially "a map" of your world view. As Bowker and Star (2000) puts it:

This mutual process of constructing and shaping differences through classification systems is crucial in anyone's conceptualization of reality; it is the core of much taxonomic anthropology. (p. 230)

The way such a project is situated (Haraway 1988) is therefore key, and understanding the use of the database is essential when considering tagging and hierarchy, which we have expanded upon in our method. You could therefore either view our hierarchy and tagging as a "snap shot of time", on how we viewed each text at the moment of creation, or you could argue that a database is a continuously developing tree, which changes, extracts, and expands depending on users, situations and time.

Inside the tagging itself, a pattern also emerges, as we have visualized, since some categories never come without the other. For instance, if a text is to be tagged with "Facebook", it also must involve "Social Media", since Facebook is a social media.

If you dive deeper into this, you would also find other patterns, such as "algorithms" often go together with "Facebook", since Facebook revolves so much around algorithms. So you could argue for a restructuring of

our “tags”, based on the correlation between them, which you can see below, where green indicates that the tags have been used together, and red indicates that they have not.

	Algorithms	Big Data	Business	Datafication	Dataveillance	Digitization	Facebook	Social Media	Marketing	Machine Learning	Politics	Cognitive	Perception	Art	Curating	Feminism	Open Data	Metadata
Algorithms	X																	
Big Data		X																
Business			X															
Datafication				X														
Dataveillance					X													
Digitization						X												
Facebook							X											
Social Media								X										
Marketing									X									
Machine Learning										X								
Politics											X							
Cognitive												X						
Perception													X					
Art														X				
Curating															X			
Feminism																X		
Open Data																	X	
Metadata																		X

You can say that we in a way have tried to create our own “system”, as Foucault (2007) describes was done with plants, dividing them into kingdom, genera and species. This way of categorizing makes each text able to be placed without having to be described. We do however stray away from this, by assigning it different tags with different lineages, which does prove troublesome, as the whole idea of “the system” is to precisely label each instance. The question could then be asked, whether “the system” has a use case in regards to texts (which are read differently depending on situations), or whether we just have not succeeded in creating a large nor precise enough tree to fulfill “the system’s” purpose?

## Assignment 4

### Level 1

#### Store

Through assignment 3, where we classified a library, it was made apparent to me, that we were not in agreement in regards to hierarchy and meaning of tags

But if we as a small group of 3 were in such disagreement, what about the bigger picture, with all the classifications going on?

Stores are only an example, but you could also think macro level, where it becomes apparent, that not everything seems to fit in this convenient box called "classification".

This ties to the case with my assignment 1, where I digitalized plants, and also assigned them tags. Some of the tags I chose, I did not know could be tied to the plant before investigating. This underlines how situated knowledge is (Donna Harraway 1988), and how different tags would resonate with some, but not others, hence making it vital, to both understand the current and future need of the data set.

#### Plants

In assignment 1 I investigated how it was possible to "digitize" physical plants, or create them into data rather, as seen on the picture. This further led me thinking, what is not data then? If even our plants are data, what cannot be?

I came to the conclusion, that everything can be seen as data, and not just one dataset. As Gitelman & Jackson (2013) describes, raw data is an oxymoron. This must then mean, that all data is situated, meaning that every single one of us have a completely different way of making knowledge out of not just data, but physical objects as well.

Take for instance these plants. If I simply removed the image, what was then left of the plant? Would they not still be digitized? How many except for botanist could instantly make meaning of it? This underlines the importance of thinking about one's dataset, and facing one's prejudices.

#### Humans

When examining data, such as this picture, you are left wondering what data really represents. Same goes for every data set. Take for example my example from assignment 2 and 3, is this data really reality, or just a representation of something from my point of view? Am I really me in Facebook's dataset, or am I left being 1's and 0's. All that seems left, is the viewers interpretation, of something that once was me.

## Time

Time is relative, as a certain gentleman once said. This is true, not only in relation to physical objects, but moreso with knowledge. Donna Harraway (1988) call it situated knowledge, and though this also means, that knowledge is largely subjective, it would also seem as if knowledge were situated into a certain time, making it crucial to understand the purpose and timeframe of certain knowledge, as it can easily expire. Entire branches of achademic studies has dwelled into this understanding, the hermeneutics, understanding the importance of not only unravelling the content of the text, but more so the time, the situation, around it, in order to fully grasp what resides within the text.

## Level 2

### Advertisements

Personalized targeting has become a real buzz word inside marketing departments over the last few decades. In reality (and simplified), personalized targeting is nothing but curated data, used to persuade users into different things.

This is what is referred to as Surveillance Capitalism (Zuboff 2015), where companies use their large quantities of data, to nudge their customers into buying their products.

Through the lens of Surveillance Capitalism, it is made obvious, why our digital self is so important to understand, since it is the baseline of what we encounter online. At the moment you can render it rather harmless, but you would not have to think far into the future, before you realize what future these data hoarding companies could create.

### Cellphone

Through my 2nd assignment, I investigated what my data on Facebook said about me, or rather, I made a "Data Selfie".

This consisted of me scripting through all my Facebook messages, in order to see when I send the most. I mapped them as seen on the left.

The result led me to see all kinds of events, which caused a change in my behavior. This made me think as to the reason I could see these changes.

Could I see them because I knew what has happened, and hence why the pattern changed, or were a similar or another motive hidden somewhere deep into the algorithmic depths of Facebook? Who has the power to see, or maybe even foresee such events. As Thylstrup, N. (2018) describes, this is an immense power. Many people do not have the capability to see through these algorithms, and even more are not aware of them. This is the true power struggle happening on not only Facebook, but the shift towards IT as a whole.



This is only made more apparent, as Shoshanna Zuboff (2015) explains, as tech companies have realized, that it is far easier to change people's behaviors, rather than predicting it.

### Surveillance

Surveillance has become a large part of our society over the years, as technology has eased the act of surveilling citizens.

In Denmark it is law, that you are obliged to make it clear to the public, that they are being monitored, through this sign for instance.

This is however not only possible by actively monitoring anymore, as more and more companies has shifted into the "surveillance economy", making them increase the amount of data they collect from their users. This could be done through cookies, where companies strive to make their "cookie policy" as complex and lengthy as possible, deeming the user incapable of actually understanding what they are clicking yes or no to.

It is however only one of many of the differences between digital and physical, though they are still obliged to inform. The issue could however be, that it could pay off to simply just ignore this, as the instance has been with major companies such as Google in the past.

A lawsuit in Arizona was filed, as Google had continued to store data of their users through Google Maps, even though they specifically had opted for the setting, to not send data to Google (Vaas, Lisa 2020).

### Bicycle Rental

More and more things are getting digitalized, and an increasing amount of mundane tasks now require the use of a cellphone.

Besides bicycles, I could mention paying at the grocerie's (COOP app), transferring money (MobilePay), paying for parking ticket (various parking apps), and the list continues.

Common for all these things are, that they are possible to do without the use of a smartphone, but require a vast amount of time. Of course these changes includes positives, since it often times are a lot faster with the smartphone, than the conventional way.

However, data are easier collected this way, which will only increase personalized targeting. This happens for two reasons, since first of all, the companies monitoring you get more data, making you get more personalized ads. Secondly, the sheer amount of data about consumers, will inevitably breed more targeting, since they increase this capital.

This is described by Shoshanna Zuboff (2015), and according to her, this will only increase, as companies try to expand their sources of data.

## Curatorial Statement

It appears to me, that we are all having a split personality. No, not on the mental level, but rather being different selves in different worlds, or platforms rather. We have become what I would describe as data subjects.

Though I will not rant about whether this is for the better or worse, I will however question: "Who am I in these different worlds and situations?"

One can understand, that global organizations has developed algorithms in order to understand who I am, but this only seems to create more unanswered questions. As to describing who I am, I will resort to Cooley (1902): "I am not what I think I am and I am not what you think I am; I am what I think you think I am" (Korgen & White 2011).

This quote meaning, that not only what I think is important, but my relation in the situation seems equally important, as to describing who I am. Actually, us humans, when interacting with both others and platforms, echo-locate in a way (Markham 2017), meaning I will try to identify with the content I face on the platform, which behind the scene has been created by the algorithm. The issue becomes, when I am no longer treated in a way, that I can identify with, creating a need for me to change my line of thoughts or my actions. This would seem quite easy in a normal setting, but proves more difficult facing a mere algorithm. How would I know how to change my behavior, in order to being treated the way I seem fit, if the only response I get is hard to decrypt and out of sync?

This question is only going to become more urgent, as the Surveillance Capitalism (Zuboff 2015) increase its span, and more sources are created for data generation and exploitation. A few of the different sources has been made apparent through the categories in this statement, and some of the concerns I have in regards to this phenomena as well. What I am arguing today is not that algorithms are stupid nor evil, on the far opposite in fact, they seem quite smart. I am merely arguing for a change in the way we perceive data, not that data is untruthful, but that it is to be taking into account the situation, and hereby goals, of which the data are created in, the expiration date of the data, and last but not least, the opportunity for the individual to freely interact and understand their digital self. Shoshana Zuboffs' Surveillance Capitalism also ties nicely together with John Cheney-Lippold (2017) explanation of dividuals, where our digital selves (dividual) controls how we are going to perceive the world around us. As Cheney-Lippold and Zuboff states, these data can (and will) be used to manipulate individuals in favor of the companies controlling the data. This is very much in line with how I have created this curatorial statement, where I guide you through my process through a pre-determined path. And truthfully, you can variate from it, but not everyone will. So not only is the data capable of changing who I perceive me, but it can also change my behavior.

My question to you then: How long will it be, before the dividual becomes the individual?

## Conclusion

It is made evident throughout the portfolio, that it *does* matter how you process and communicate data. Things change when they become digital, and whether or not they are a correct representation of the individual will remain individual.

However, the importance of thought behind this labeling and processing is unquestionable, and calls for more opacity from the large corporations who not only hold these massive amounts of data of their consumers, but even process it, and sell it to the highest bidder.

In order for the individual not to be forgotten, changes are ought to change. The consumers need not only to be made more aware of this categorization, but demand more influence of the whole process, in order to not be represented correctly online, and worse; becoming the misrepresentation.

## Appendix

Accessible at <https://vindvind.eu/curatingdata/> (Google Chrome is recommended) under "Other Assignments"

## Bibliography

- Acker, Amelia, og Adam Kreisberg. *Social Media Data Archives in an API-Driven World*, 2020.
- Bowker, Geoffrey C., og Susan Leigh Star. *Sorting Things Out: Classification and Its Consequences*. 1. MIT Press, 2000.
- Cheney-Lippold, John. *We Are Data: Algorithms and the Making of Our Digital Selves*. NEW YORK UNIVERSITY PRESS, 2017.
- Connor, Michael. "Curating Online Exhibitions", 13. maj 2020. <http://rhizome.org/editorial/2020/may/13/curating-online-exhibitions-pt-1/>.
- Foucault, Michel. *The Order of Things: An Archaeology of the Human Sciences*, 2007.
- Gitelman, Lisa, og Virginia Jackson. *"Raw Data" Is an Ozymoron*. MIT Press, 2013.
- Haraway, Donna. *Feminist Studies*, 1988.
- Korgen, Kathleen Odell, og Jonathan M. White. *The Engaged Sociologist: Connecting the Classroom to the Community*. Pine Forge Press, 2011.
- Markham, Annette. "Echo-locating the digital self", 2017.
- Thylstrup, Nanna Bonde. *The Politics of Mass Digitization*. MIT Press, 2018.
- Turkle, Sherry. *The Things That Matter*, 2007.
- Vaas, Lisa. "Google sued by Arizona for tracking users' locations in spite of settings", 29. maj 2020. <https://nakedsecurity.sophos.com/2020/05/29/google-sued-by-arizona-for-tracking-users-locations-in-spite-of-settings/>.
- Zuboff, Shosana. "Big other: surveillance capitalism and the prospects of an information civilization", 2015.

## Icons from the webpage

Data stream: <https://www.google.com/url?sa=i&url=https%3A%2F%2Fdzone.com%2Farticles%2Fdata-streaming-in-osgi-r7-applications-with-osgi-r&psig=AOvVaw3zh3Kcqe5U8HjZ57HD2KOb&ust=1608230771259000&source=images&cd=vfe&ved=0CAIQjRxqFwoTCMjVsteU0-OCFQAAAAAdAAAAABAE>

Facebook logo: <https://1000logos.net/facebook-logo/>

Eye of Sauron: <https://www.emaze.com/@AIZOWTFR>

Reload: <https://findicons.com/icon/457357/reload>

Questionmark: <https://pixabay.com/vectors/question-mark-circle-blue-button-38595/>

Buttons: <https://www.clickminded.com/button-generator/>

Excel icon: <https://i.pinimg.com/originals/de/cd/de/decdde1594f5073dca58c89cd502bef0.png>

PDF icon: <https://lh3.googleusercontent.com/proxy/EYH6svngxn-Ra3u9SZo9XYOqW6LJ6YtxnAyZ3Na-YkYQ8B497sOjUdIQPt6ZHVBvkGXP7E0jYaFnuYgBA7dPQntm7iXhxtBsOfK7YZbJoZ7E0TxX8F008kfAmtE>

Image icon: <https://i.stack.imgur.com/0dKeS.png>

Python icon: [https://cdn4.iconfinder.com/data/icons/logos-and-brands/512/267\\_Python\\_logo-512.png](https://cdn4.iconfinder.com/data/icons/logos-and-brands/512/267_Python_logo-512.png)

Text file icon: [https://lh3.googleusercontent.com/proxy/Fzliv7YBT-id-wQdrYOHGHcl18FbfU\\_Kkkccz6PA9UrVHPmIx9uVDGV\\_d\\_2TYmJ266bEFEloPeLPp-MUQ6olrHxnc\\_yG0CgJr4NZHMGL8jrymglyXXfkjZ6AU\\_pVH03nAsXvl8brjCrYISUS82NAIAwg](https://lh3.googleusercontent.com/proxy/Fzliv7YBT-id-wQdrYOHGHcl18FbfU_Kkkccz6PA9UrVHPmIx9uVDGV_d_2TYmJ266bEFEloPeLPp-MUQ6olrHxnc_yG0CgJr4NZHMGL8jrymglyXXfkjZ6AU_pVH03nAsXvl8brjCrYISUS82NAIAwg)

Crosshair: [https://upload.wikimedia.org/wikipedia/commons/thumb/9/95/Crosshairs\\_Red.svg/1200px-Crosshairs\\_Red.svg.png](https://upload.wikimedia.org/wikipedia/commons/thumb/9/95/Crosshairs_Red.svg/1200px-Crosshairs_Red.svg.png)

Radio waves:

<https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.needpix.com%2Fphoto%2F171212%2Fradio-waves-yellow-broadcasting-sending-wlan-wifi-wireless-lan-electromagnetic-field&psig=AOv-Vaw2yGrHMpLA-IHaq4syPRGow&ust=1608237594530000&source=images&cd=vfe&ved=0CAIQjRxqFwoT-CNiN2I2u0-OCFQAAAAAdAAAAABAD>

Bicycle app: <https://cdn.donkey.bike/wp-content/uploads/2017/05/16120921/then-and-now-app-768x432.png>