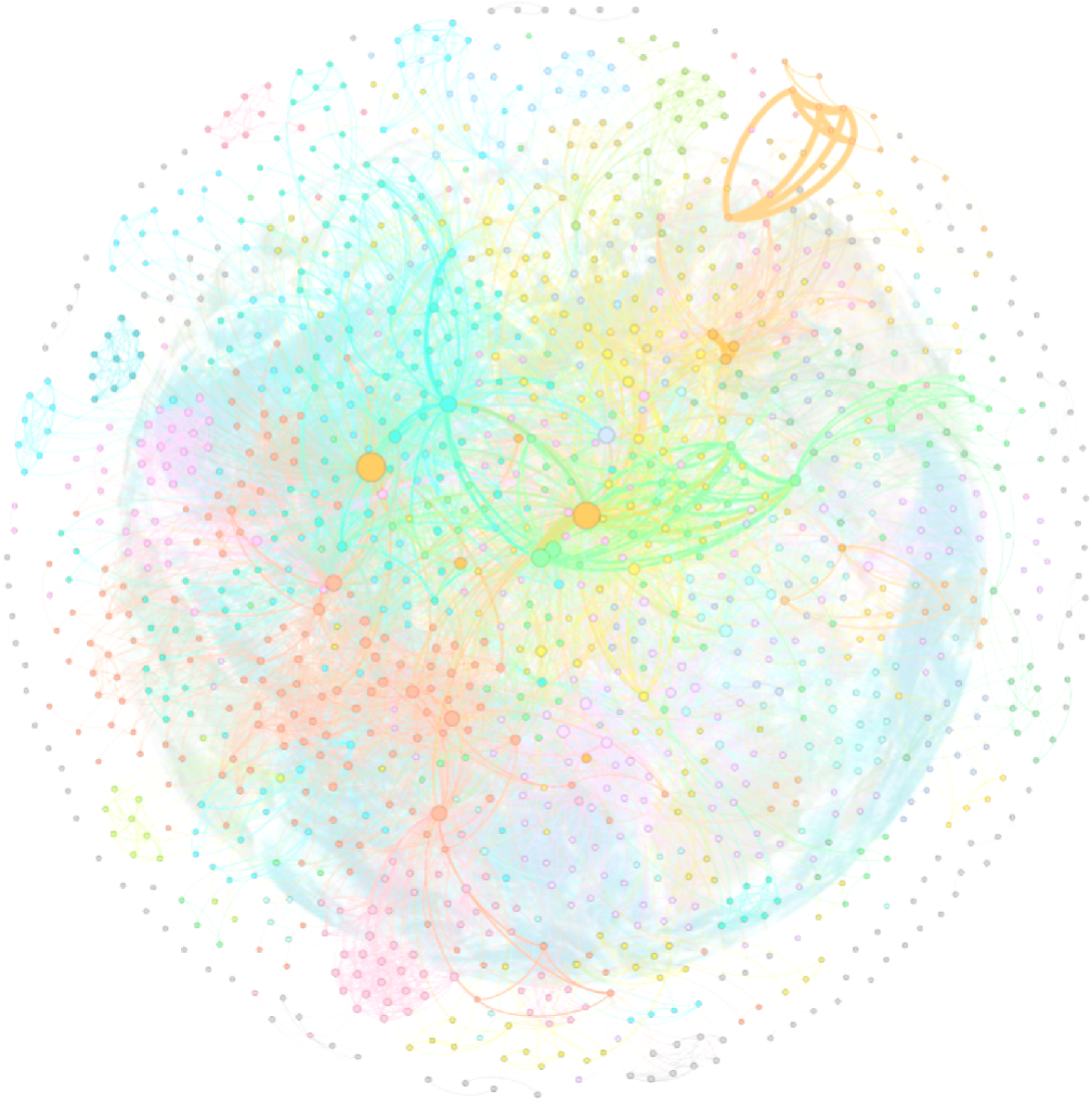
Unboxing open government data

A valuable something - somewhere - out there - to someone

Master Thesis by Anders Grundtvig



Aalborg University, Copenhagen
Techno-Anthropology

Title: Unboxing open government data; *a valuable something - somewhere - out there - to someone*

Name	Study Number	Signature
Anders Grundtvig	20126460	

Place of study: Aalborg University, Copenhagen

Education: Techno-Anthropology

Semester and type of report: 10th semester - Master Thesis

Supervisor: Anders Koed Madsen

Date of handing in the report: 07.06.2019

The size of the report: 80 Normal pages

The report is not confidential

AALBORG UNIVERSITET

Abstract

Introduktion: I takt med den fremadstormende digitalisering i Danmark, genererer det offentlige i stigende grad digitale data. Offentlige myndigheder, ministerier og kommuner genererer, i deres daglige administration og arbejde, data om det samfund vi alle lever i. Dette er data om transport, byggetilladelser, adresser og CVR-registeret. Men det er også data om cykelstier, e-cigaret regler og børnehave lukketider. Alt dette offentligt genererede data er nu begyndt at blive distribueret gratis og åbent, under betegnelsen "åbne offentlige data".

Initiativet med åbne offentlige data bygger på, at det offentlige alligevel generer alt denne data, så hvorfor ikke lade alle bruge det?
Rapporter peger på at åbne offentlige data er værdifuldt for virksomheder at få fingrene i, men på trods af store forhåbninger, er åbne offentlige data stadig ikke blevet taget så massivt i brug som det offentlige havde håbet på.

Nærværende speciale ønsker at undersøge de barrierer der ligger til grund for denne manglende adaption af åbne offentlige data. Specialet argumenterer for at denne slags undersøgelse skal være data-drevne og data-nære for at kunne overkomme brugernes barrierer med åbne offentlige data.

Specialets problemformulering er derfor: *Hvad er de primære socio-tekniske barrierer for åbne offentlige data - i den ikke-statslige brug. Hvordan kan data-drevne visuelle repræsentationer og fortolkninger af åbne offentlige data åbne nye diskussioner og indsigter om brugertilpasning og barrierer for åbne offentlige data*"

Nærværende speciale undersøger denne problemformulering ved at fokusere på *brugerinddragelse* og *data-nærhed*.

Det empiriske grundlag for specialet bygger på to initiativer. Først et halvt års arbejde med åbne offentlige data fra sommeren 2018 til vinteren 2018 som med-facilitator på initiativet UrbanTech Challenge, der var organiseret af Partnerskabet for Åbne Offentlige Data.
Dernæst et halvt års egenhændigt arbejde fra vinteren 2018 til sommeren 2019 med organisering og facilitering af fire kvalitative data-drevne workshops, samt 4 stakeholder interviews.

Det teoretiske grundlag for specialet bygger på Akrich og kollegaers (2002a og 2002b) begreb om socio-teknisk innovation med fokus på *interessement modellen,* der foreskriver at innovation forekommer ved at interessere involveret og samle humane og non-humane aktører om det innoverede.

Det metodiske grundlag for Specialet bygger på en participatorisk design tradition, hvor brugerne bliver inviteret ind i maskinrummet og er med-designere. Denne tradition bliver udbygget med et eksplicit design fokus på åbne offentlige data, som her bliver medieret gennem data visualiseringer. Disse visualiseringer bliver metodisk om-konfigureret til et *designobjekt* i den kvalitativ participatorisk design workshop sammen med brugere af åbne offentlige data. Disse design objekter har til formål at stimulere en datadrevet data nær samtale ved at invitere brugeren til at interagere med faktisk data, inspireret af participatorisk *data* design.

Specialets bidrager med ideer til data-drevne, kvalitative empiriske situationer. Med data visualiseringer og design omgivelser der stimulerer *brugspotentialer ved serendipitet* (Björneborn 2017). Dette udtrykker hvordan man opdager, udforsker og tillære ny viden. Specialet bidrager dermed til hvordan brugerne kan oplever en ny måde at tilgå åbne offentlige data.

Det problematisk ved distribueringen og åbningen af åbne offentlige data er omfanget af offentlige myndigheder der udsteder data. Åbne offentlige data bliver udstedt på over 84 forskellige data platforme; med forskellige tilgange til adgang, brug og formidling - hvad jeg kalder "platformenes logik".

Specialet bidrager afslutningsvis med at lokalisere fem barrierer for åbne offentlige data som alle leder over i én stor strukturel barriere. De fem barriere er: Manglende tillid til data ressourcen. Høje standarder for tekniske produkter. Manglende fokus på brugerne i tekniske produkter. Etiske bekymringer og særligt en indbydende og sansevækkende data konfrontation.

I alle disse barrierer eksisterer, hvad jeg kalder en iboende "restprodukt tilgang" til åbne offentlige data, der medfører en ikke-orienteret og ukonkret udvikling og tilgang til åbne offentlige data.

Specialet afslutter med at foreslå, hvad jeg kalder, en "kurateret data tilgang" hvor åbne offentlige data på et strukturelt og lavpraktisk niveau skal udstilles, administreres, bygges og designes med brugerne i fokus, og som produktet frem for restproduktet.

Table of contents

1. Introduction - Welcome to the ecstatic urbanism

Imagine a city any city can do, as long as the imagined city swarms with ecstatic busy urbanism. A city with a great activity that in harmony with its surroundings, moves organically around in between solid structures that reach for the sky, waves in and out of public transportation and electric cars. Imagine a city free of wandering garbage, without traffic jams and with citizens fully informed of public affairs and the development of the city. Imagine this city, and then start to look around. If you look close enough, you will start wondering how this urban organism is organized, and instead of seeing *the lady in red*, you might start inspecting the means of movement and liveliness; *the infrastructure*. Roads, bike lanes, traffic lights, public transportations, parking lots and electricity cables are all means for movement. But also, what has come to be known as the social infrastructure appears, the design of the city square, the small green spots, parks, benches, open spaces, and open buildings are all means for social interaction and liveliness. Again you ask yourself, how are these infrastructures organized? To answer this question, we have to enter a more abstract space, where all the concrete and all the existing physical infrastructures are turned into zeros and ones as computational data points. Imagine that you can see the same city as observed before - now just through its present as digital data points. An entire urban infrastructure, social and concrete, is now in its new form interconnected and interactive as digital data. Data that are making the city accessible by machines and people confidently seated in chairs behind their screens. Imagine a parking lot in digital communication with cars. An automatically emptied garbage bin. A city that silently communicates and structures itself while social activity and movement swarm around without recognizing it. Welcome to the coveted (and yet highly imagined) frictionless Smart City.

1.1 Governmental generated data

The government, the authorities, and the municipalities are with urban administrations, public city planning and general daily administration, every day producing and dealing with increasing amounts of government-related digital data. Data about the urban infrastructure, but also about processes and procedures related to the affairs of the given authority. The Danish Business Authority is generating company information data. The Ministry of Higher Education and Science are making data with insights about the educational system. There are being

generated data about e-cigarette rules, data about restaurant sanitation smileys and data about fishing, agriculture, buildings, citizenship, churches, and the list continues.
There have been, and still are a political engagement of opening this government-generated data up for the wider public, leaving it on open data platforms in the reachable distance for everyone to download and use - all free of charge. This accessible governmental generated data travels around acknowledged as *open government data*[1] (abbreviated; open gov data). The literal concepts that encapsulate this technology might externally be the immediate concepts of "open access", "governmental generated" and its format as "digital data points" - but as we are going to dive deeper into the concept of open gov data, we are also opening the lid to a technology installed in a highly political culture sphere. We are going to grasp the technology not solely as a technological means of zeros and ones, but in its natural social habitat as a socio-technical phenomenon, a technology embedded in a culture of specific understandings and recited narratives.

But the hopefully desired goals for open gov data are still standing in the shade of realities. Realities might reveal that great data stories travel faster than actual data implementations and use as we are going to see. But it seems like it is not only big data prospectors and government narcissism that drives the hopeful advocating for open gov data. Worldwide initiatives, especially in Great Britain and the US, are shedding a glorious light upon open gov data initiatives (Opengovpartnership.org., 2019), reports from big consultancy and open gov data initiatives around the world are promoting this data source as a goldmine for companies (Europa.eu, 2018) and open gov data projects in academia encourage new possible transparency-, livelihoods-, and public affairs projects (Davies and Bawa, 2012 and Morelli et al., 2019). The critical concept in Smart Cities is *data*, in this case, open gov data. *Open gov data* are going to be the subject for scrutinization in this Master Thesis.

The Danish government is by multiple political initiatives trying to enhance and aid the usage and enlargement of open gov data, all in the scope of the informal, and bit drowsy, slogan; *"we are generating the data anyway, why not try to have more people use it?"*. The desired usage is, in this sense, not an internal governmental use - but an external non-governmental use.

[1] Translated from Danish, "Åbne offentlige data"

1.2 Open governmental initiatives

One of the leading figures in open gov data initiatives is the relatively newly formed Partnership for Open Government Data[2] (abbreviated; The Partnership). The Partnership is a collective of four significant public instances, relevant for this project is The Danish Business Authority and Opendata.dk that I have experienced as the motivating factors for the development and growth of open gov data in Denmark.
One primary object for The Partnership in their consistent journey has transformed away from a general focus and instead to find and highlight particular open gov data success stories. This political interest for individual cases leaves us at a Partnership organized project[3] where I am in the role of facilitator and in the position, given by the participants, as an open gov data expert are situated together with a participating company that in great eager to understand open gov data has compressed his eyebrows. We are debating how he can adapt open gov data in his business concept.

"Okay," he says, "I understand the key concepts of open government data, that this data that is created in the work of administering the government are now shared with open access for us, companies, to make use of. But, I still do not understand what data we are talking about and where it is". "The thing is," I answer, "that there are such big amounts of data that I cannot in details explain to you what is out there, and even where it is, is difficult for me to elaborate. The data is linked to its public creators, and the creators are all located in different places with different prospects for making this data". He looks at me, nods a bit, and says, "but you want me to use this data, right?". "Yes," I reply. "But how can I use data, that I don't know about and are located on platforms I do not know about?". I hesitate a moment, desperately trying to look like someone with the correct answer, that just are struggling to find the right words to articulate this answer. But fails. "Look, I am the facilitator of this project. We have to direct your question to the data owners[4]. But we have completed a list of open gov data examples that are posted on the website that I can show you, and I also have some other sites that might be relevant for your business. We are also hosting a data owner/data user meetup next week that you can join. And of course, you can always try to search the web for 'open government data' and surf a bit around, the data is indeed out there

[2] Author translation from; "Partnerskabet for åbne offentlige data."
[3] https://challenges.dk/en/challenge/urbantech-challenge
[4] The public employees that are issuing and managing open gov data.

somewhere". "Year okay, I get it, there is a lot of data out there" he replies, "but taking the time and effort into account I think I am going to continue developing my concept on the data I already know about, namely CVR data and address data. I think this data is interesting enough."[5]
Some days later we, the facilitators, are invited to the manager table of The Danish Business Authority department of open gov data where we together with data owners from the open gov data source "plandata" are discussing the prospect of using the plandata data source. The plandata data owners start out presenting their data by telling the story of the department. With a great distance to any particular dataset, we are instead discussing issue specific governmental protocols and procedures. The data owners add, "Remember, each of these building permission datasets is created locally by the municipality issuing it. Most of them are stored in non-readable formats for machines like an unstructured pdf and are created to deal with a particular concern in a particular case. In the near future, we might be able to convert these datasets into machine-readable and human understandable formats, but we need first to show the urgency for this to happen". As we are discussing the structure of this specific data source, it is suddenly getting more and more clear for me that open gov data is a complicated thing to grasp. What at first glance seems uncomplicated to encapsulate as one homogeneous thing, is now starting to reveal its real face as a heterogeneous unstructured concept that is only being held together by the mutual term; open gov data.[6]

This Master Thesis tries to exploit and experiment with the usability (or the lack of the same) of open gov data in Denmark. There are still a lot of problems to face and solve before this data source fully can be recognized as a value-producer-machine, and before it is widely adopted by non-governmental users, as illustrated above.
One of the most fundamental issues when trying to delimiting open gov data, as one united field is that this data source has multiple faces, multiple forms, and various reasons for existing. Another issue is that the potential users do not know how to find the data or the data platforms since the data is spread across multiple platforms that even makes it impossible for the government itself, to understand the comprehensiveness of this data source. I have, therefore, in this Master Thesis, created a variety of data visualizations that can direct, facilitate, and open new discussions about open gov data. I have in this Master Thesis intensively been working on these visualizations to make various data entry alternatives that could provide the user with insights and overview of the present open gov data in new forms.

[5] Author field note, UrbanTech Challenge, 2018
[6] Author field note, Plandata, 2018

The Government initiated open gov data project tend to be centered either about a technological data near frame or a social frame with the users in focus without data proximity, but not social/technical combined. One primary object for this project is, therefore, to combine social initiatives with data near prospects. The way that this combination usually is taking place is, e.g., by inviting potential data users like companies to an arrangement together with the so-called data owners, the people administrating the data in the government. These arrangements are seeking new ground for public innovative where they, e.g., are facilitated as a casual dialogue cafe around high tables or as speed-dating with the companies on one side of the table and the data-owner on the seat across. These playgrounds for public innovation do have data in focus since the whole idea of inviting the data-owner is to introduce the companies for open gov data, but the data tend only to exists in the spoken language where no data platforms are visited, and no datasets are being opened. This Master Thesis seeks to experiment with new playgrounds for combining data and users in the same setting, in the scope of being in the field as an active part of it, while intervening it from inside.
This intention is building on the foundation of general interest and belief in the socio-technical scope of understanding the world, where social material and technical material are interwoven and are acting concerning each other. Akrich and colleagues (2002a and 2002b) are arguing that innovation processes are the art of simultaneously treating the social and the technical in the same space of development. They say that social transformation and technical development are two sides of the same innovation and argue that innovative adoption is through technical and social adaptation.

With all this taking into account, this Master Thesis strives to obtain and produce user knowledge and user insights to enrich and enlarge discussions about barriers for open gov data in Denmark. Provided with time and resources for contemplation I have, in the catchphrase of "*slowing down reasoning*," explored this field of open gov data with the following research question.

1.3 Research Question

What are the primary socio-technical barriers for open government data to be adopted by non-governmental users? How can data-driven visual representations and interpretations open new discussions and insights about user adaptations and barriers of open government data?

The overall theme in this question is a vast focus on open gov data narrowed down to its barriers for broader usage. The question expects barriers to exists since the author on an earlier project have outlined the existence of barriers. The next important aspect of the question is the socio-technical. This is one of the critical elements of Techno-Anthropology methodologies and Science and Technology Studies (STS) concerns and leads us to another critical epistemology element namely the data-driven and the visual representations as means for user insights. Combined fields of computational science and social science are fostering new possible data interpretations to appear and become an active part in qualitative interview situations.

2 Ontological consideration

Open gov data, as a broad concept, is ontologically positioned in a diverse frame of existing. Open gov data is a natural means of making the city smart by digitizing the properties of the city, to allow the city to communicate and thereby become *smart* (Gaur et al., 2015). This smart city delusion of city properties controlled and managed on the internet, also referred to as "the Internet of Things" is both feared and endorsed, but as the Smart City researcher Martin Brynskov acknowledges in a radio show[7] smart cities are already here, and most of us live in them right now.

The data processed in this Master Thesis reveals the same conclusion, city trees, city toilets, the roads and so forth are all available as digital data points, as digital blueprints of the city, on open gov data platforms ready to be downloaded and used by everyone - but in practice mostly by bigger companies and the government itself.

Another ontological position of open gov data is within the frame of the "open" prefix movement, (Chalmers, 2012) accompanied with *open* data, *open* source, *open* innovation, and crowdsourcing as other examples in the movement (Davis and Bawa, 2012). The open-movement is an ideology that believes in transparency and open access from a democratic socialist perspective (Smith et al., 2011). OpenDefinition.org (Open Group, 2018) is defining the movement by; *"Open data and content can be freely used, modified, and shared by anyone for any purpose"* (Open Group, 2018). From an ontological perspective, open gov data is in this project positioned between the smart city and the open movement. This Master Thesis has an immense focus on discussing users experiences related to the openness element, while at the same time fostering new usabilities of the properties of the smart city prospects.

Morelli and colleagues (2019), drawing their attention on open data to the democratizing of the data usage to perceive the data as an opportunity for innovation in smaller communities rather than large corporations.

Morelli and colleagues (2019) argue, with the term "community of users," that open data should be considered as *commons*, that by-design invites everyone to use it. The emerging question is

[7]https://www.elektronista.dk/digital-kultur/elektronista-radio/elektronista-shownotes-351-smart-cities-er-ansvarlige-cities/ . Visited 05.06.2019.

if the ideology of the open data commons here are investigating the technological development of open gov data or just are educating a public.

Von Hippel (1986) propose that technological development needs experienced users, what he terms as *lead users.* Lead users are acknowledged as crucial factors in the starting phase of technology development to forecast the potential market. *"Since lead users are familiar with conditions which lie in the future for most others, they can serve as a need-forecasting laboratory for marketing research."*(Von Hippel, 1986)

Lead users are in the case of open gov data often acknowledged as the data-driven tech companies. The fundamental problem with data-driven companies as leads users is that they are supposed to drive open gov data in the favorable direction for the technology, but this favorable development might only serve the lead users and not the broader public in general. But as Von Hippel (1986) argues, most users are not suitable to understand the technology product and its prospects simply because of missing technical skills. *"most potential users will not have the real-world experience needed to problem solve and provide accurate data to inquiring market researchers"*(Von Hippel, 1986). In an interview with Seismonaut[8], we discussed their experience with inclusive and democratic projects where they openly invite everyone to participate. These projects end up with enlightened participants, but also with no further inputs for the development of open gov data and with no concrete afterward established projects. I had similar experiences with participants without a concrete project at hand, that resulted in unconcrete fluffy discussions instead of the desired data near discussions. So, in the journey for tangible socio-technical discussions, the data mature with particular cases was in favor of the extended democratic, open data commons.

[8] A private consultancy company that has been arranging and facilitating open gov data dialogue meeting and conversations cafes for the government.

3 Theory - innovative actors

"The socio-technical analysis underlines that the movement of adoption is a movement of adaptation." (Akrich et al., 2002b)

When dealing with open gov data, terms like *innovation* tends to appear - in two different appearances. Innovation in open gov data is either scoped internally, where open gov data is the object for innovation, or externally, where open gov data is used to foster innovation.
Open gov data is created and distributed, with the governmental expectations, to support innovation by one user refining and distributed it to another user. This can be seen as the external innovation that open gov data are supposed to support innovative aspects out in the world. Internally, open gov data in itself, as a technology, is the subject for further development and thereby innovative progress.

3.1 The key to success

Madeleine Akrich is, together with Bruno Latour and Michael Callon in a two-part paper, "*The key to success in innovation*", part one; "The art of interestment" (Akrich et al., 2002a) and part two, "The art of choosing good spokespersons" (Akrich et al., 2002b), discussing innovation in the extension of an Actor-Network Theory (ANT) inspired socio-technical analysis. They are asking the overall question; how can unforeseen success and failure be explained? And to answer this question, they are directing their attention for explanations to a field; they have been co-developing, namely ANT.

ANT is an academic discipline in the field of STS developed by Latour, Callon, and Law. Where other disciplines are obsessed with *explanations*, ANT instead focuses on empirical *descriptions*. ANT scholars approach the world from a constructivist viewpoint, arguing that, e.g., knowledge is a *social product* stabilized in a heterogeneous network of actors rather than a product of objective science (Law, 2992). The mobilizations of actors in a network of alliances are here dissolving the distinction between macro and micro, agency and structure and are instead asking to treat materials, people, machines, ideas, discourses, equally as a network "*as interactional effects rather than primitive causes.*" (Law, 1992:389)

Akrich and colleagues (2002a) are in order to discuss the importance of different aspects in the innovation process, introducing two terms to separate the process, the specific innovation ability to spread independently; *The diffusion model,* and the process of building actor-networks; *the model of interessement*[9]. The diffusion model proposes that the innovative force is situated inherently in the properties of the innovated, where it is the passive innovative thing-in-itself that holds the prospects of success. The *model of interessement* proposes that innovation is a mean of an active surrounding environment and holds that the interessement is involving an active network-actor where adaptation is recognized as identical with adoption. Akrich and colleagues (2002b) are using the concepts, *a movement of adaptation* and *a movement of adoption* as a direct dependent of the other, in the *active* actor model of interessement, where innovative development is a direct means of provoking the environment to foster adaptation and thereby adoption.
Akrich and colleagues (2002a and 2002b) are throughout the two texts holding the promise to conclude how to make successful innovation. The texts are, of course, not revealing a simple formula for successful innovation. Such formula would have made the texts become an international bestseller and would have made life way more uncomplicated for the writers and developers around the world. Instead of proposing a simplified innovation formula, Akrich and colleagues (2002a and 2002b) are with their socio-technical analysis, complexifying innovation processes with their intensively focus on the interessements.

"Innovation is the art of interesting an increasing number of allies who will make you stronger and stronger." (Akrich et al., 2002a)

This quote is the closest we are getting to a formula for innovation; aggregation of interest with alliances that increasingly make your product stronger and stronger. This is where a socio-technical analysis is being highlighted as an important method to grasp the aggregation of interest. A socio-technical analysis is an explicit focus on the interrelation between the social "material" and the technical "material" with the ambition not to separate the two. Akrich and colleagues (2002b) argue that innovation is to develop and transform both the product and nature around it. One of the ambitions with the socio-technical analysis it to be situated in the

[9] "Interessement" is a merge of French and English words, invented by Callon (1989) where he in relation to the study of power describes the word as: "a series of processes by which the researchers sought to lock the other actors into the roles that had been proposed for them in that programme" (Callon, 1989). In this Master Thesis, the term is used as: "the art of interesting an increasing number of allies" (Akrich et al., 2002a).

development situation where both the technology and the surrounding environment can be analyzed with the ambition to realize the interconnectedness between the two.

Akrich and colleagues (2002a and 2002b) argue that social feelings and technical feasibility both are a part of innovation, which is why they say that human factors and technical factors cannot be separated. Within this argument co-exists their central message, adoption is an adaptation. The argument holds that both the technical material and the surroundings have to shape for the innovation successfully to be adopted.

Akrich and colleagues (2002a) are in the first part obsessed with interessement as actor-networks that they directly and indirectly in the text favor over the diffusion model. But in the second part, they are turning their interest in the innovative role back on the individual in favor of the broader interessement, namely on *strategically spokespersons*.

Akrich and colleagues (2002b) are highlighting the importance of actor-networks by building up a marked for it. The aspect of non-human objects that actively are acting in the actor-network is one of the main theoretically tricks in ANT (Law, 1992). This allows non-human actors to play a prominent role in the network, but this also draws attention to the spokespersons of the given technology, because who is presenting themselves as the representatives that provides the technology with a voice in the debate? Callon (1984) is showing how the "winner" of being the spokesperson for a given object can influence the surrounding debate in favor of the spokesperson. Innovative spokespersons, as presented by Akrich and colleagues (2002a and 2002b), are in similarity to Callon (1984) a design process strategy to drive the development and discourses in the desired direction. Akrich and colleagues (2002b) round their text up by highlighting that the manageable aspects of innovative processes are to be an active actor and spokesperson with the role making confrontation against confronting discourses in the game of governing strategic decisions.

This way of conceptualizing development in a socio-technical analysis with focus on *interessement* and *spokespersons* and with an active focus on negotiations, alliances, non-human actors and the concepts of adaptations as adoption are going to be subjects for further analysis when we again are turning our attention towards open gov data as an innovative process - *in the making*.

4 Method - participatory data approach

We are in this chapter further going to build the very foundation for the further Master Thesis to rest on. I am going to introduce and discuss crucial concepts and exciting thoughts that bright people have developed and expressed. These concepts are in this chapter combined and orchestrated together for me to use this orchestration later on as references and methodological support in later arguments.

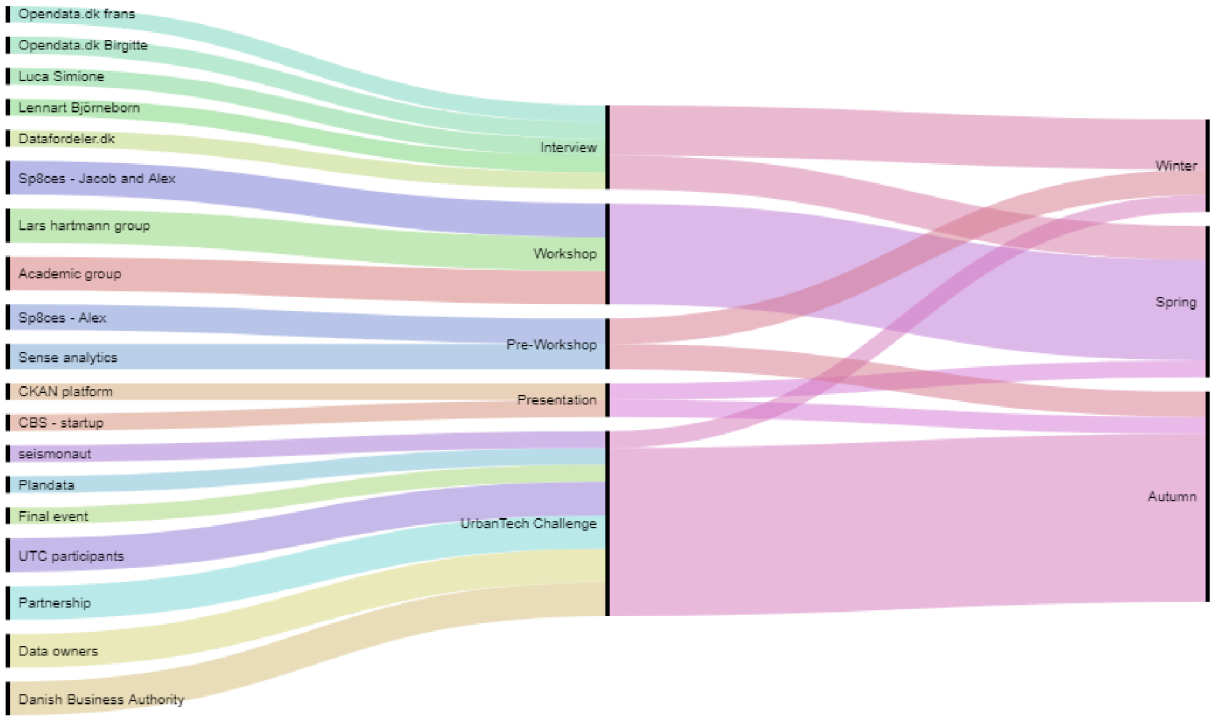
We are here starting with concepts of qualitative methods and are afterward continuing into design objects, positioning in the field and the end a suggestion for an interventionist approach. But first, we are introduced for the empirical situations that have shaped the project.

4.1 Empirical work and study design

The foundation of this Master Thesis rests on a broad array of empirical situations and theoretical conceptualizations. I am here outlining the empirical work that has been the backbone of the project.

Inspired by serendipitous (Björneborn, 2008) encounters, I am here with a visual exhibition showing the process of the empirical work. Visualization 1[10] shows, from right, i) a column with the names of the most important actor that I have been in contact with during My Master Thesis. ii) a column with the given situation, and iii) the time of the year. Please note that the properties are structured according to support the visual interpretations (straight lines) rather than alphabetical or chronological[11]. The lines are sized according to effort and time I used with the actor[12].

[10] See appendix 1 for the same data in a tabular structure
[11] This, e.g., influence the order of the right column with time, where autumn is actually before winter but are here visualized as after.
[12] This number is just an estimate - and most interesting to design the visualization.



Visualization 1 - Empirical field work

4.1.1 UrbanTech Challenge

I was from August 2018 until January 2019 at a scheduled academic Internship[13] in the company Greentech Challenge[14]. Greentech Challenge facilitated the UrbanTech Challenge[15] that was organized by The Partnership, and managed by The Danish Business Authority[16]. UrbanTech Challenge was about nudging companies to use open gov data by rewarding the best concept with attractive prizes. Participating companies competed in the prize by developing a business concept that smartly included open gov data to enhance their business plan. Participants would get involved in workshops[17], get in contact with data owners and general always contact Greentech Challenge to receive help and finally participate in a grand final event[18]. The facilitation of UrbanTech Challenge positioned me deeply involved with the open gov data and involved with the promotion and communication of open gov data. During the facilitating and as

[13] https://www.sadp.aau.dk/digitalAssets/358/358232_msc-i-teknoantropologi—kbh-_21122017.pdf p. 22
[14] https://www.greentechchallenge.eu/
[15] https://challenges.dk/en/challenge/urbantech-challenge
[16] https://erhvervsstyrelsen.dk/
[17] https://www.eventbrite.com/e/urbantech-challenge-workshop-tickets-50475081327
[18] https://bloxhub.org/event/urban-tech-challenge-the-final-event-in-danish/

the role as an open gov data expert, got me entangled with data concerns and conceptualizations both by the Business Authority, data owners, and users of open gov data; the participants of UrbanTech Challenge. By being involved with the participants' problems and concerns with their process of finding and encountering data, I experienced data frustrations and barriers on first hand.

4.1.2 Interview

In order to gather deeper and qualitative broader understanding of the concept of open gov data, I conducted four interviews. I interviewed the consultancy company Seismonaut[19] about their experience with facilitating open gov data projects. I talked with Frans from opendata.dk[20] about his perception of the users of open gov data. I discussed data platforms with datafordeler.dk[21] and I discussed ideology with Luca Simeone from the Open4Citizen[22] project. I discussed serendipity theory with Lennart Björneborn[23] and I am going (after this project is handed in) to discuss data contents with opendata.dk and Smart Aarhus[24].

4.1.3 Observation

After being fascinated by the material structure and communication at libraries I spend about one month at Rentemestervej library[25], observing how the users interacted with the space and how they found and encountered material. I also discussed the decision-making process of selecting material with the staff.

4.1.4 Data Workshops

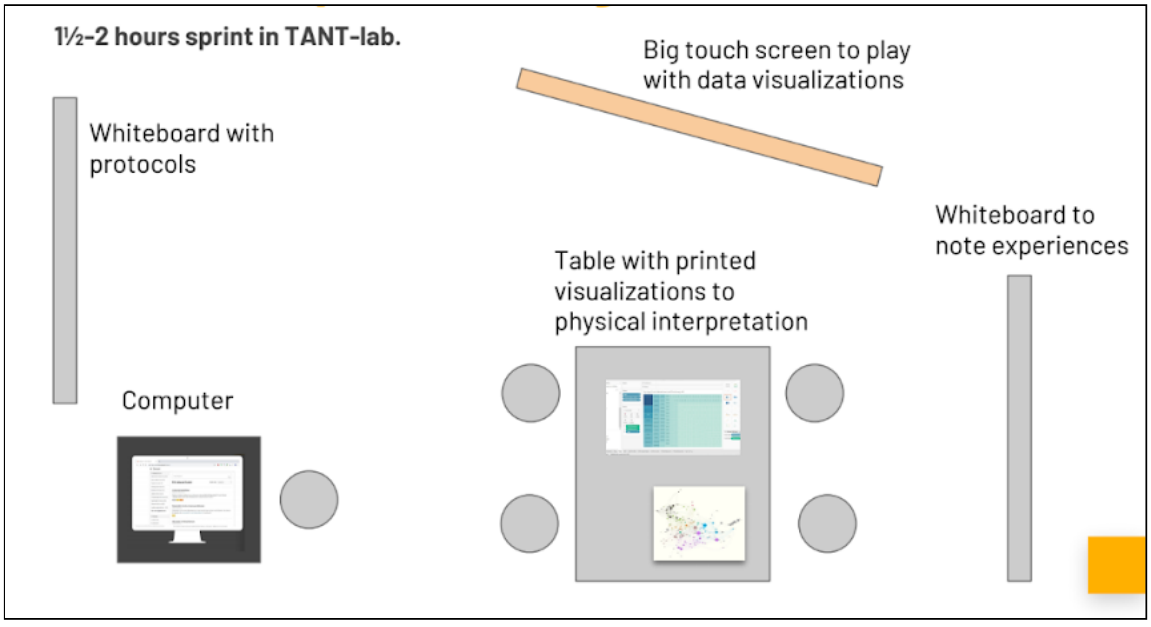
In the termination of UrbanTech Challenge, I experienced and located several issues with open gov data that had been present during the challenge. These experiences were outlined and discussed. I developed an interview concept of which I anticipated would enhance the data entry and data proximity for the users. This resulted in several data visualizations. I experimented with the format and the visual expression and conducted pre-workshops to test

[19] https://seismonaut.com/
[20] http://www.opendata.dk/
[21] https://datafordeler.dk/
[22] http://open4citizens.eu/
[23] https://informationsstudier.ku.dk/ansatte/?pure=da/persons/451979
[24] https://www.smartaarhus.eu/projects/open-data-dk
[25] https://bibliotek.kk.dk/biblioteker/rentemestervej/om

the setup with Alex from the company Sp8ces and the company Sense Analytics. After this *graphic ideation* process, I conducted three data workshops in TANT-lab at Aalborg University.

The workshops were designed to facilitate data proximity though data visualizations and concrete dataset and empower the participants to encounter the data. The idea was that this facilitation and empowerment elicited data-driven discussion and hands-on experience, rather than the usual data distant approach. To obtain this desired objective, we exerted a data visualizations interpretation exercise. The participants and I systematically interacted with each visualization and decoded if it contained interesting content. If this was the case, we found it and opened the related dataset.

In order to build this data proximity at the workshop, I ended up spending a lot of time explaining and showing the landscape of open gov data throughout the data visualizations and the data platforms. The workshop went for one and a half hours. The workshops were facilitated at TANT-lab[26] because of the big touch screen and to create a design space exclusively for investigating open gov data. Here the setup is sketched, visualization 2:



Visualization 2 - Workshop setting, TANT-lab

[26] https://www.tantlab.aau.dk/

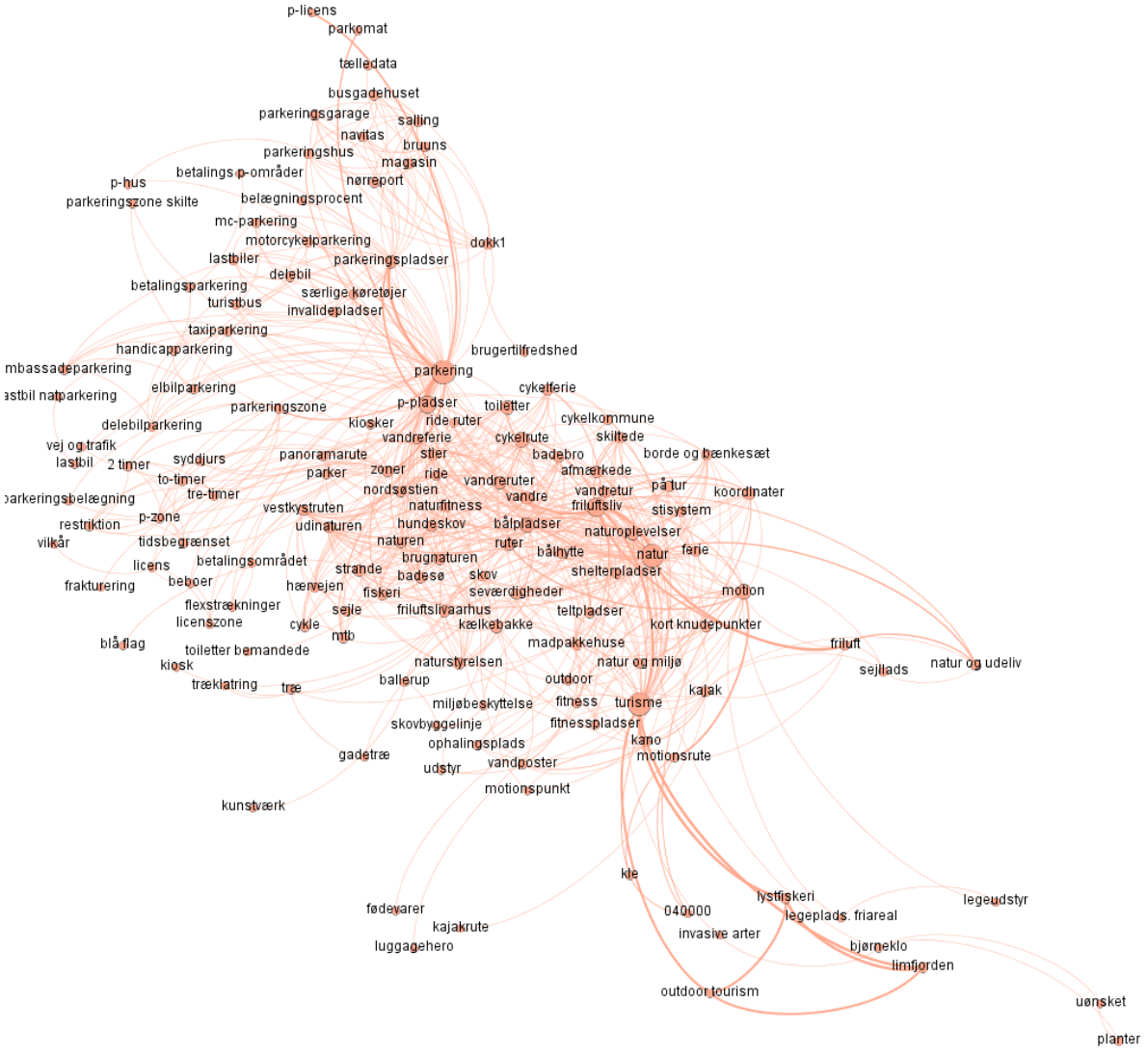
4.2 The design space

Here we are, inside a very square room behind a floor-to-ceiling transparent glass wall, situated among summing expensive computers, whiteboards, a red couch, a long table with to many chairs, virtual reality sensors, roll-up banners and a huge touch screen that beside being an alluring piece of hardware also functions as a blockade for the mess behind it. These things constitute the physical and non-human technical space known as The Techno-Anthropological Laboratorium, TANT-lab. I have been giving the credibility to animate the lab and turn it into the designated design space, a design playground, as it was fabricated for (Abildgaard et al., 2017)*. The long table was decorated with printed open gov data related materials, e.g., a catalog produced by the Danish Environment Portal, the winning concept of UrbanTech challenge and a list of open gov data portals and descriptions from UrbanTech Challenge. The whiteboard is prepared with two squares drawn on it, one saying 'desired data,' the other one saying 'new ideas.' On the big touch screen, the first page of the introductory presentation with the data visualization is present. I am situated in front of my computer that is controlling the big touch screen. I have several taps and programs opened on the computer now visible on the screen; a presentation with 35 slides, the portal on opendata.dk, the open gov data report by Rigsrevisionen, Carto website, "R" with the two prepared datasets - ready to ask a new question to the dataset, Gephi with manipulatable graph visualizations, and Tableau with the two interactive visualizations. I have decorated myself with a shirt and black jeans to manifest, mostly for myself, that this is the event that I have been preparing for.*[27]

This is the physical and technical set up for open gov data workshops. I am later going to describe the mentioned datasets and how they were produced. I am likewise also then going to *explain* how the mentioned visualizations were generated with the datasets. Just hold the tension we will get there. The workshops were prepared and designed in association with the design space methodology discussion that we in the following are going to explore. But without revealing too much already, I can unveil an example of a data visualization, to give some impressions of what we are talking about, see visualization 3.

[27] Author fieldnote, Workshop, 2019

The workshops were intended to facilitate a design space with the primary focus on empowering the technical material into the social material and the social material into the technical material; namely, to facilitate a socio-technical design situation (Akrich et al., 2002a and 2002b)[28]. The *technical material* is here the actual concrete open gov data, and the *social material* is the discourses, preconceptions, and surrounding setting of open gov data pursued by the user. The open gov data are for this situation translated into data visualizations as an attempt to create a manageable and informative and at the same time tangible *Information*



Visualization 3 - Data visualizations graph, opendata.dk, no1

[28] We are also going to elaborate further on this later

World Map that the users could interact with. The data visualizations are planned to be the starting access points for the open gov data for the users in the qualitative situation. The actual concrete open gov data, e.g., found on the portal at opendata.dk was then planned to be an access point for arrival during the workshops. The social .material presented by the users were supposed to be focusing on hands-on experiences by actual consumers of open gov data. The intended idea was to foster concrete social material by designing the workshop with an immense focus on concrete user situations in a reflective space that allowed *new worlds* to appear. The open gov data visualizations are here framed in the scope of being a design object, *design game/graphics elicitation/elicitation device*[29], that mediates the technical material into the social material and vise versa by provoking a relating between the two. This aspect is operationationalized in a qualitative situation in TANT-lab at the described design space together with the users and along with the design objects; the data visualizations. This is the topic for the next chapter to introduce and elaborate on.

Let me elaborate on the concept of *the design object* and the concepts of designing a design space and the other concept mentioned above, by sketching out a literature review about *design objects* in *qualitative situations*. This review includes qualitative methods, design spaces, participatory design traditions, design objects as elicitation devices and ends in a field of digital data and data visualizations within the field of Digital Methods. All these concepts are here going to be introduced with the intended destination in mind to combine and use them at the above-outlined design space.
But first, the literature review - of the field *participatory data design* here enhanced with theories of design objects.

4.3 Participatory data design "*objects*" - a review

Qualitative-method driven disciplines within the field of social science share content analysis, interviews, and fieldwork as primary methodologies for knowledge production. Both fieldwork and interviewing are, from an anthropological perspective, aiming to obtain deep insights by interacting with the field of study and to obtain what Gertz (1973) famously called *thick descriptions*. Roughly speaking, fieldworkers gains insight by participating in the actual field and study the *doings* in the field. Interviewers seek insights by facilitating a situation detached from the doing, empowered by communication and question-driven insights (Sunstein and

[29] hold the tension; we are very close to elaborate on these methods.

Chiseri-Strater, 2011). Traditionally qualitative research has been focusing on face-to-face situations, but newer examples are experimenting with the traditions.

Internet research, especially the digital-prefixed ethnology like *net*-ethnography, *cyber*-ethnography, and *digital*-anthropology, has shown how to conduct internet-based field work. Miller and Slater (2000) with their notions of online and offline *being* in the field. Further Boellstorff (2008) and Nardi (2010) explore how "proper" fieldwork can be translated into the fieldwork of the internet. Hine (2000) discusses how traditional methods are ignoring certain aspects of technology-cultures and argues that Internet-based field work can capture a level of detail and produce thick descriptions in specific fields that traditional physical-driven fieldwork are not prepared to - since the notion of the *space* is reconfigured with the digital.

Qualitative methods have by social scientists been modified and enhanced by implementing technology into the situation (Murthy, 2013). Murthy (2013) is discussing how technology like the field recorder, mobile phones, and the searchable internet are turning classic ethnography into technology-ethnology. Hine (2000) is discussing how internet-based meetings and interviews are changing the scope of anthropology.

People born within the era of the internet and private computing, like myself, might find it obscure to recognize these everyday technologies, like online communication and mobile phones, as classifiers for technology-driven research but technologies are becoming an active part of the knowledge production also in the more traditional fields. The classic interview situation is indeed changed and challenged when it comes to online interviews, isolated from a common physician space. Fieldwork might experience a technology changed practice due to smartphone development and private computing. This technology shift is an exciting development for qualitative social science.

This Master Thesis is academically framed in the field of STS and Digital Methods as proposed by Rogers (2013), and Birkbak and Munk (2017) with an explicit focus on digital data points.

Scholars in the field of traditional ethnography and cultural anthropology are sometime referred to an opponent of integrating *Big Data* in their research (Curran, 2013). Traditional ethnography and cultural anthropology are referred to, as grasping the Big Data concept with the infamous *what* instead of the more traditional academic *why* (Mayer-Schönberger and Cukier, 2013) as a threat to their qualitative work. Curran (2013) is advocating for a move in ethnography and cultural anthropology and proposes that these disciplines should learn from the business

adoption of big data. He argues that both anthropology and the business world are interested in the same insight in people's lives. He proposes a new field, Big Ethnographic Data (Curran, 2013). Curran (2013) is aiming for an ethnographic future that brings the big data world into the ethnographic field and develops a shared interpretative approach for research. He does not explain in further details how this goal is supposed to be fulfilled or operationalized and does not dive into any methodology discussions for such an approach, this is being scrutinized in the following were I am investigating how interviewing and interview situations can adapt data-driven research tools by introducing data as a visual stimulus in the interview setting.

4.3.1 Design the design space

Participatory design is a qualitative approach that aims to include stakeholder in the process of designing solutions (Robertson and Simonsen, 2012). A methodological cornerstone in participatory design is the focus on the design space itself. Robertson and Simonsen (2012) define participatory design as a mutual learning process between participants and the collective in a *'reflection-inaction"* space. Brand (2006) further argues that designing the design process is as important as designing the artifact. This argument holds an explicitly focus on the design space as a mean for innovation and therefore, also a vital part of this Master Thesis. We are later on exploring the design space as we are exploring how design objects are introduced in the reflection-in-action space. Reflection-in-action is interesting from an STS perspective concerning the act of being present in the *science in the making* space. This allows the researcher to become an integrated part of the field and understands and criticize it from inside, what Bruno Latour (2005) acknowledges as critical proximity in essential favor of distance. Birkbak and colleagues (2015) further argue that the critical proximity perception avoids premature references of the field and implify that the researchers are implicated in the issue that is researched. This study is highly inspired by this proximity approach, which will hopefully be appearing in the thesis where I am framing it as *data proximity*.

4.3.2 Design objects

A crucial methodological concept in this Master Thesis is the exploration of using data visualizations and actual open gov data as an elicitation device in the qualitative discussion space. I have previously experienced that open gov data discussions and workshop tend to be extremely distant from actual data. By facilitating a participatory design space with the users, I am aiming to explore new aspects of open gov data usage and data concerns. This methodology

is further motivated by the socio-technical analysis (Akrich et al., 2013) where the technical material (open gov data) and the social material (the user, use, and ideas) are shaping each other.

The act of introducing materialities, design objects, or other kinds of visual stimulus in the interview situation is a widely used method in the Scandinavian participatory design tradition. Examples include mock-ups, (Brandt, 2007 and Rosenqvist and Heimdal, 2011) design games, (Brandt, 2006), prototypes (Robertson and Simonsen, 2012) and design probes (Mattelmäki, 2006).
This is also present in a broad array of other qualitative research fields, eg. with Information World Mapping (Greyson, 2013) participatory *data* design (Jensen et al., 2017) art-based methods (Bagnoli, 2009), information horizons (Sonnenwald et al., 2001) elicitation devices (Abildgaard, 2018) and graphic elicitation (Crilly et al., 2006 and Bagnoli, 2009).

Brandt (2006) discusses the use of *design games* in interview situations and argues that game materials and the gaming-situation can enhance the reflection space with the participants. She argues that it facilitates new types of reflections and improve the act of *re-seeing* the issue. These game materials are to Brand (2006:64) recognized as *things-to-think-with.* Greyson (2013) highlights the advantages of what he calls *Information World Mapping.* This is a project that animates the participants to take part in a creative communication process of drawing their relations (world maps) to specific situations, places, and operations. The drawing aims to help the informants to articulate and highlight their relationships to a given subject. Information World Mapping is trying to shift the interview conversation focus from the *source*, into the *practice* by introducing objects for communication.
Earlier anthropologists have also argued that visual objects could augment traditional interview methods. One example could be John Collier (1957:858-859) that argued that pictures, as a research aid, can provoke longer and more comprehensive interviews and kept renewing the energy in the interview setting.
Crilly and colleague (2006) discuss how the interview situation is transformed when introducing stimulus to the interview table for interpretations, what they call *graphic elicitation*. The concept is to introduce graphics that encourage or provoke discussions and *"break the frame of the subjects normal view"* (Harper, 2002:20, in Crilly et al., 2006). Crilly and colleague (2006) facilitate the graphic elicitations interview setting by introducing a relatively simple flow

diagram into the situation. This situation strives to facilitate a position that stimulates the participants to interpret and change the frame of the researched subject by commenting on the flow diagram. The flow diagrams are then changed according to the feedback from the participants. This iteration is repeated over and over to come closer to a diagram showing an actual structure of the study. This process is called *graphic ideation* (Crilly et al., 2006). Compared to the Scandinavian participatory design tradition, graphic elicitation can here be recognized as related to prototypes (Robertson and Simonsen, 2012) since it is a representation of an actual world and an actual product. It is also related to design games since it aims to foster the same type of user elicitation and similar activity/gaming processes in the interview situation. One significant difference between the two design materials approaches "design games" and "graphic elicitation" is how the materials/games represent the lived world of the participant what we could call *world maps.* The design game approach deals with abstract concepts to be interpreted and rearranged (Brandt, 2006) and the graphics elicitation approach strives to make representative world map visualizations.
Graphic elicitation is according to Crilly and colleagues (2006), quickly achieving credibility by the participants since it is showing the world maps of the participants themselves. The graphics elicitation objects are meant to animate inspirations and creativity. But they can end up being persuaded by the participants as explaining the field rather than leading to the intended reflective thinking, exploration. This problem is caused by means of considering the objects as being an accurate reflection of the world map rather than being a temporary model for interpretations and device for the purpose of elicitation (Crilly et al., 2016:359).

Abildgaard (2018) is offering a phenomenological aspect with her use of elicitation objects in qualitative interview situations. Abildgaard (2018) is introducing material artifacts with a historical perspective related to it, what she acknowledged as *a materially oriented qualitative interview*. She argues that these artifacts are eliciting another type of memory that allows the informants to be bodily remembered and reflect on their world maps. This approach aims to help informants recalling thoughts and feelings. Staying on the notion of "recalling familiar aspects," Guggenheim proposes a method that is in opposition to familiarity and recall. Guggenheim and colleagues (2013) strive to produce knowledge claims that do not reproduce a common ground and common knowledge. They seek to introduce objects, materials, stimulus, and toys to facilitate a space where new world maps can appear. Guggenheim and colleagues (not afraid of mis-credibility) are taking games and interpretations to an extreme in their

interview situation with the goal of advancing the qualitative output. They argue that their approach is beyond the scope of common ground and does not correspond to usual academic systems. This means that their findings are not simply translatable and commensurable with previous findings and understandings. Guggenheim and colleagues (2018) are further discussing how interviews are supposed to complicate situations and open up the possibility to articulate new worlds. This is according to Guggenheim and colleagues (2013, 2018) perceived by facilitating new interview situations, like described above, and by asking new types of interview questions. The facilitation of *"modes of questioning or interacting which produces realities"* as they describe it are what Guggenheim and colleagues (2018) call *incubation,* a space for new worlds to appear. This explicit focus on new worlds is motivated by criticism on traditional uniform questionnaires and interview questions and argues that these traditional methods are simply inviting people without previous opinions about the topic to invent opinions designed to that survey. (Guggenheim and colleagues 2018). This facilitation of incubation and new worlds as Guggenheim and colleagues (2013 and 2018) are advocating are directly comparable with an interventionist approach as we in the following further are going to explore.

4.4 Intervening and experimenting - with design objects

"*We hope to contribute to the creation of a grammar that allows for both nuanced reflection on and exciting experiments with what it means to 'intervene' as an STS researcher.*" (Zuiderent-Jerak and Jensen, 2009)

Zuiderent-Jerak and Jensen (2009) are with this quote building some of the foundations for the interventionist approach in STS to expand. Zuiderent-Jerak (2015) is further investigating the interventionist approaches in social science and suggest a study design with experiments inspired by natural science studies. Salganik (2016) is with an exclusive focus on internet-based social science, also suggesting an experiment-driven research approach.

Marres (2017) advocates *re-combined* methods that can be used interactively with the fieldwork almost like Hogan and colleagues (2007) but also concerning Madsen and Munk (2018) and Munk and Colleagues (forthcoming). These re-combined methods are also of interest in Taylor and Horst (2013) that discuss how ethnographers can compose thicker descriptions of the narrow field by understanding an *aggregate picture*, as they write, using different tools and digital data to understand this picture. Blok and colleagues (2017) are on Curran (2013) proposal for Big

Ethnographic Data, experimenting to bring, as they write, "*transactional and ethnographic data into close conversation",* in order to show that data can be big and broad but can also be deep and thick, where they argue that their combined data approach augments the thick ethnographic description.

Data sprints, as proposed by Munk and colleagues (2018) and Venturini and his overlapping colleagues (2018) are an example of STS research to facilitate participatory data design and critical proximity in an intensive workshop situation. Data sprints and especially participatory data design is attractive in behalf of the operationalization of the interventionist approach with design objects in focus.
In an introductive video (tantlab, 2018) Anders Munk and Torben Elgaard explains how data sprints can be used as a method for early participant engagement in the design process. Munk and colleagues (forthcoming) further show how to involve the participants in an interventionist data process. Munk and colleagues (forthcoming) are reframing the setting of a data sprint with a simple interventionist trick. By merely removing labels on a data visualization, they are changing the knowledge production focus away from the data and instead into the participant. The involved participant was told to engage themselves with the visualization and by doing so, repurposing the visualization into a design object - that fostered new discussions among the participants. In another project, Madsen and Munk (2019) are reconfiguring the social media platform, Facebook, into a specific participatory data design project. They are engaging relevant involved actors with a Facebook page designed and created as a part of a grand vision day in Aalborg municipality. Madsen and Munk (2019) changed the production of knowledge claims away from the data and reconfigured the data to a design object to be used as a graphic elicitation with the participants. The visualization did in this behalf exist partly as a design object and partly as an information source that the participant could make claims *from,* rather than an autocratic source of knowledge.

4.5 Method sum up

This methodological introduction of graphic eliciting devices and qualitative situating are the main methodological foundations for this Master Thesis. I am especially building upon Munk and colleagues and Madsen and Munk (2019) interventionist conceptualization of reconfiguring the data visualization. This approach is in this Master Thesis explicitly being combined with the notion of graphic elicitation and design of the design space.

We are further in the analysis going to use and animate these above-outlined participatory design concepts to obtain knowledge and produce knowledge claims about open gov data.

4.6 Back inside the design space

Jacob and Alex from sp8ces are now both standing up in front of the big screen. I have more or less forced them into this vertical position, but they seemed satisfied as they in enjoyment are investigating the bombastic piece of hardware that the big touch screen constitutes. The atmosphere was calm, relaxed, and neither of us was, after one hour of discussing and talking, no longer intimidated by small periods of silence. We had been conversing and discussing how open gov data could benefit their company sp8ces. We had written down concepts on the whiteboard that Alex and Jacob found exciting and was now more explicitly focusing on the graph visualizations on the big screen in order to look for new interesting concepts for further attention. We went through the ten graph visualizations each containing clusters of connected topics, one by one. Alex and Jacob are equipped with a pen to mark on the touch screen. I am equipped with a whiteboard marker. They highlight interesting topics on the screen. I record the topics on the whiteboard simultaneously. 'How about "this"' Alex starts, 'maybe that could be something to look deeper into?'. Jacob, not convinced, replies, 'how should that aid our product?'. 'I am not completely sure, but we did earlier talked about focusing more on traveling business people, and maybe we have a connecting here?'. 'I get your point,' Jacob replies, 'but in that light, maybe this small cluster over here is more interesting?'. Alex and Jacob, engaged in interacting with visualizations, are discussing back and forth about how to aid the existing concepts and also new concept rises during their interacting and cooperating with the visualizations.[30]

We are here presented for a small passage of the qualitative situations in TANT-lab together with Alex and Jacob from the company Sp8ces. The data visualization are activated as a things-to-think-with as Brandt (2006) describes design games, where the visualizations are mediating the contents of open gov data into a tangible *thing* that here in the design space are used as an object to ask new questions. The animative role of the visualizations in this situation holds many similarities with design games. There is no visible *gaming* aspect in the design situation if games are considered as competition, but there are certain gaming aspects when framed as stimulations and idea advancing and generating.

[30] Author fieldnotes, Sp8ces, 2019

The visualizations perceived in the light of Crilly and colleagues (2006) *graphic elicitation* rather than as a Brandt (2006) *design game,* opens the importance of the visualizations as contents and information contained and not just as playable objects. The playful motivator for the users interacting with these visualizations is the explicit value of obtaining insights about reality while playing with the visualizations. This aspect is crucial in these workshops since it provides the design space with a sense of reality that allows the users to become familiar with open gov data while obtaining data proximity. This data proximity concern was interesting to follow as Alex and Jacob equipped with the paint-less pen worked themselves through the visualizations one by one. By being two participants, they naturally talked together, articulating their concerns and ideas as they interacted with the visualizations. The visualizations were here clearly visually eliciting the conversation and were here aiding to maintain the communication on the technical material; open gov data as social materials, (concerns, ideas, frustrations) emerged along the interacting with the visualizations.

You, as the reader, might have noticed how I continuously declare what the visualizations in-themselves are cable of doing and eliciting. The visualizations, here assigned as eliciting value in themselves, are of course not sparking any new discussions and reflect new worlds to appear just by being present. The visualizations must be animated together with the entire design space. The big touch screen is mediating the interpretative playfulness atmosphere as well and likewise, is the room also influencing the eliciting and influencing *the reflecting space*. As we saw in the field note above was I likewise a part of the data visualization animating, as I asked them to take place in front of the big screen.
Another interesting aspect to discuss is how much the visualizations foster new worlds maps to appear as earlier questioned. Crilly and colleagues (2006) show how their design object, the flow diagram, was limiting some of their participants to create new ideas, new worlds since the flow diagram was perceived as being a fact rather than as a design device for visual eliciting of new aspects and thoughts. The technical material is, in that case, determine what kind of outcome that is possible. This limited possible outcome of the technical determinism can also be examined on the social material determinism. One aspect of social material determinism is how open-minded and welcoming atmosphere is for new worlds to appear - and new worlds are not necessarily a product of excitement, champagne, and prestigious Danish design furniture. During the Danish Business Authority organized UrbanTech Challenge was all the participants extremely excited and open-minded about open gov data. In the processes of serving and

aiding the participants in the challenges, co-work with the authority, facilitate workshops and facilitating the final event show, nobody questioned the structure of open gov data or related themselves to their critical sides. I would argue that UrbanTech Challenge indeed was a success in relation to the ambitions and objectives for the challenge. But my point is here that UrbanTech Challenge did not encourage a *reflective space* to appear.

We have now, inside this design space at TANT-lab been introduced to the participatory data design and the design objects that have been introduced to the space as elicitations devices for discussions and as a guide to enhance the participants encounter with open gov data. But what is this object? How does it look? I am in the following showing the object, but first, we are going to inside the machinery to explore how they were produced - and the considerations they were built in continuation of, namely the *concerns*.

5 The technical part of the project

5.1 Data visualization, a product of concerns

I have throughout the beginning of my personal open gov data journey since the beginning of the summer 2018 encountered problems related to technical and social aspects of open gov data. As described in the introduction, the most frequently raised concern when entering the messy field of open gov data is uncertainty about *contents*, with users asking *"..but what data is out there?"*. The next two issues experienced during UrbanTech Challenge was related to data proximity among the participants and the data-drivenness of the initiatives. The three significant concerns thereby are; *the concern of data contents*, *the concern of the users data-proximity*, and *the concern of data-driven initiatives*. These three concerns lead me in the direction of developing open gov data visualization.

The visualization that I planned to create should include and cover the following aspects and concerns. Communicate some overview of the available data. Provide a simplified version of the data landscape. Show the comprehensiveness of the data. Provide a Venturini and Latour (2010) inspired data zooming ability that allows grasping the data from a micro and macro level simultaneously. But also produce a visualization that fosters open-ended creative interpretations for interview situations and visualizations that visually are interesting enough to encourage interpretations.

It is essential to keep in mind, as described in the method part that this project and these visualizations are intended to be a part of an experimental and intervening methodology where the visualizations can be acknowledged in the scope of design objects or prototypes rather than final products ready for the market. The visualizations should facilitate overviews, and immediate insights that situated together with the users could open new qualitative discussion through data proximity.
I am in the following elaborating on the developing of the data visualization that is going to be used as design objects. And to build data visualizations, we need data.

5.2 Where does the data come from?

I have in this Master Thesis intensively been working with two datasets. The datasets choice builds on the effort to representing as much open gov data as possible to show a full picture of the open gov data landscape.
The two chosen datasets that I am working within this Master Thesis each represent a public-body administration issuing level, the first dataset, *opendata.dk,* contains data on a municipality issuing level and the second, *Rigsrevisionen,* contains data on a ministry issuing level.

5.2.1 Opendata.dk dataset

The opendata.dk dataset, as the name might reveal, is harvested from the opendata.dk open data portal[31]. The data on this portal is an association of, often geospatial, data about selected municipalities in Denmark. Examples of this data could be *bike lanes*, *electric car charges, shelters* and so forth, often data about the physical cities - but more on the specific contents later. The data portal on Opendata.dk is built on the open source platform CKAN[32] which gives it a similar interface as other portals build on CKAN platforms. When data owners in the municipalities are issuing data on a CKAN platform, they are asked by the platform to answer several questions about the dataset being issued. The platform is, e.g., asking, *how often is this dataset being updated? Who is maintaining it? What is the name of the dataset? How would you describe the dataset?* And also a question that turned out to be particularly interesting for this Master Thesis; *Describe the content of the dataset with a few indicators (informally known as "tags").*

"Tags," compared to the "title" or the "description," holds an uncomplicated opportunity for a computer to read and compare datasets across the data portal. These tags are small summaries of the content of the dataset. This type of *extra* data about the actual data, like tags, titles, descriptions, IDs, and so forth, are on CKAN stored together with the actual dataset and are categorized as *metadata* (Hey and Trefethen, 2003).

[31] https://portal.opendata.dk/
[32] https://ckan.org/

Metadata is a quite interesting data source when performing data analysis since it tells something about the data in a simple format that rather simple can be compared across the portal and gives the ability to investigate the datasets in different ways and on a different level (meta levels) like time periods or maintenance patterns (Hey and Trefethen, 2003).
This Master Thesis is not concerned about the specific locations of all trees in Copenhagen (as one dataset contains) but is instead curious to investigate how this tree-dataset is related to other datasets in the portal, with the goal in mind, to locate themes and topic communities in order to produce a greater visual overview of open gov data. For this purpose, metadata is a unique data source, and especially the tags turned out to be favorable.

The CKAN platform has a powerful rest-API[33] that allows everyone to manipulate data with a variety of different procedures on the portal, including scraping data in a structured fashion. CKAN portals are designed with a web browser interface (eg. portal.opendata.dk/[34]) where everyone, simply by clicking around on the webpage, can search and make filters in regard to searching for datasets. The CKAN API allows everyone, human and non-human to execute the same actions available through the web browser interface just without visiting the webpage and without *clicking* around. The CKAN API in relation to this Master Thesis allowed me to scrape[35] metadata from all datasets available, this action is not possible through the website.
The metadata was scrapped as a tabular file in a .csv format containing all the answers from the questions that CKAN asked the data issuer, *name*, *description, tag* and so forth. The answers constitute the columns and each row represents a single dataset from the opendata.dk portal. The sixteen questions now sixteen answers one in each column.[36]

This dataset, *the opendata.dk dataset*, contains metadata from 979 datasets about municipality data, from the portal at opendata.dk.

[33] API, is a way for software to communicate. The platform is deciding what data can be accessed through the API. Data can then be requested from the platform through this API. This can automate data processes and aid a big amount of data access.
[34] https://portal.opendata.dk/
[35] Data scraping or data mining is a computational technique to extract data from a source
[36] "Organization_title, City", "organization_name", "packageid", "license_title", "num_resources", "name", "title", "url", "version", "author", "author_email", "maintainer", "maintainer_email", "metadata_created", "metadata_modified", "notes" and "tags"

5.2.2 Rigsrevisionen dataset

The Rigsrevisionen dataset is, again as the name might indicate, a dataset obtained from work completed by Rigsrevisionen[37] in regard to their comprehensive open gov data project about locating and discussing all the data that each ministry issues and holds[38]. There was no interacting with API or any coding skills required when obtained this dataset. Rigsrevisionen had, probably in respect to the *open*-part of the investigated subject; *open* gov data decided to publish the dataset that they had used to complete their report about the current open gov data issuing situation in the Danish ministries. Rigsrevisionen had assigned every ministry in Denmark to complete a list of questions that, when returned and combined with all the other ministries, constituted the dataset I ended up using in my Master Thesis. The combined ministry answers are published in a .xlsx format[39]. Each dataset is a unique row, and the questions with answers are the eight columns. The eight columns are "*Ministerium", "Nr", "ID", "Datasættets titel", "Beskrivelse af datasæt", "Dataansvarlig myndighed", "Udstillingssted"* and *"Link til datasæt".*[40]

This dataset, *the rigsrevisionen dataset*, consists of metadata from 952 datasets about ministry generated data, stored across 84 different government managed platforms.

5.3 How to make a data visualization 1-0-2

Now, with the opendata.dk and the rigsrevisionen dataset, we have our two base dataset collected[41].

The initial idea with these two datasets was to use them as means for data-driven discussions and data proximity in qualitative situations with users. But data in tabular forms with rows and

[37] A public body instance that "revises on behalf of the Folketing and strengthens responsible management for the benefit of the citizens."
[38] http://www.rigsrevisionen.dk/publikationer/2019/122018/
[39] Microsoft Excel file format containing tabular data
[40] Please note that the data issuer was not asked to answer the question of tags, and the datasets are therefore not containing "tags" as metadata. I built a small program that aided me to transform "titel" and beskrivelse" into tags. See Appendix 2.
[41] A chronological storyline would address, that this dataset collecting process initially only concerned the opendata.dk dataset, mostly because of availability. But after a few data demonstrations and data discussions with users, it stood clear for me that the project should include another broader dataset. This fostered the motivation to collect and include the rigsrevisionen dataset.

columns, containing almost 2000 rows in total, is extremely difficult to interpret and bodily understand - except for a computer.

5.3.1 General data refine

To produce useful and insightful data visualizations both for interpretation in a qualitative situation and as a representative blueprint of open gov data, general data refining was necessary to animate the datasets. The data refining, or with a more accurate and charged word, data *manipulation* process did consists of a variety of choices and decisions that influenced the results. Data are always manipulated and created for specific situations by someone and thereby always loaded with intentions and biases (Boyd and Crawford, 2012). The two datasets in this project are no exception. To support the user data interpretation, I, e.g., decided to "clean" the dataset for misspellings and to align plural endings. By applying these cleaning decisions, I reinforce the concrete intended usage of the data for this project. But if the same datasets afterward are imagined to be used to investigate problems related to the process of having multiple individuals to tag the different dataset, this "cleaned" dataset is no longer "clean" but instead, "spoiled"! And if this cleaning process is not well described the dataset is instead of being a spoiled data, straight up a cause of false results since all misspellings have been resolved. This quite situated and narrow example is here my attempt to show that data are stuffed with intentions and are build with a specific goal in mind as I have for this project.

The base datasets are going through a data manipulation process where it first is being cleaned in the coding language "R" using Rstudio[42] and then in Openrefine[43]. R is my preferred tool for data manipulations since each process step easily can be reversed and repeated. In R, the columns of the dataset are renamed and unuseful are deleted. "Tags" and "titles" are adjusted to lowercase. Openrefine is a data cleaning tool that makes it rather simple to apply in-build string (words) clustering algorithms. Openrefine can help and support detecting groups of similar words when spelled differently or when contains different endings and even looks-alike, e.g., "friluft" and "friluftsliv" is being detected by the algorithm, that I in each instance manually had to decided to merge or not (this example was merged). This process was to align the datasets and make them coherent.

[42] https://www.rstudio.com/
[43] http://openrefine.org/

The tags were in the spatial graph visualization tool Gephi[44] converted into dots, also known as *nodes* (Venturini et al., 2014). Nodes are in Gephi connected with a line, also known as an edge (Venturini et al., 2014) if the nodes were used in the same dataset. The tag "elbilparkering" is, e.g., in this instance, connected with the tag "delebilparkering" since both tags are being used in the same dataset, namely, "P-pladser."

Spatialization algorithms, build into Gephi, forces the nodes away from each other while the edges hold them together. I like to imagine an enormous expanding space where clusters of planets in solar systems are held together through gravity - but are pushed away from anything else. The "ForceAtlas 2" spatialization algorithm (Jacomy et al., 2014), creates a visual representation of the datasets, where the tags that are used in the same datasets are located closer to each other in this two-dimensional space. This creates groups of tags, also known as *clusters*.

To create a human interpretive visualization of the data, different in-build Gephi algorithm was applied. The algorithm, *modularity class*[45], detects groups in the spatialized network. These groups can then be used to colorizing the visualization and later on be used to divide the visualization into several visualizations representing different topics groups of open gov data.
The nodes were sized according to how many connections they had, known as *a degree*. The edges were sized according to the number of connections between two nodes, also known as, *weighted* edges. The modularity algorithm, in Gephi, mutates a new column in the dataset containing a number, representing the cluster that the tag now is a member of. This number was used to create clusters of topics.
Each modularity class cluster was then manually inspected to detect the content and an overall theme of the cluster. Each cluster was then renamed according to the manually identified theme.

[44] https://gephi.org/
[45] https://github.com/gephi/gephi/wiki/Modularity

5.3.2 The data visualizations

5.3.2.1 Gephi

The text above explained how Gephi was used as a computational analytic tool to create topic clusters and also how the data was visualized with nodes and edge size. The visualization was further, with interpretatively in mind, visual enlarged with labels on the nodes and aesthetically designed with curved nodes and the Fruchterman Reingold spatial layout algorithm[46], that was chosen in the final visualization. This algorithm was selected because of the existing small cluster that is not connected to the main network also known as, cluster islands (Venturini et al., 2014) that disappeared in the ForceAtlas 2 layout.

The two graph visualizations, one for the opendata.dk dataset and one for the rigsrevisionen dataset was both containing too much data to visually interpreted for human beings. Each holistic graph visualization was, therefore separated into eight detached cluster graph visualizations according to their modularity class clusters. Labels were now visual, and the graph interpretable.

The Master files of the two datasets were also used to produce another type of visualization, interactive visualization in the program Tableau.

5.3.2.2 Tableau

Tableau[47] is a drag-and-drop visualization tool that allows the researcher to connect a variety of properties from the datasets into a dashboard where these properties interactively can be investigated and visualized. This, e.g., enable the researcher to compose an attractive visual dashboard that subsequently, in an interactive manner, can be interacted with and explored. Regarding the overall topic of this Master Thesis, it must be noted that Tableau is anything but open source software.

Two visualizations were produced in Tableau both with a Science Po Medialab (Venturini and Latour, 2010) inspired micro/macro zoom in mind. The initial idea with the visualization was to equip open gov data users with a tool that allowed and encouraged her to choose topics of

[46] https://github.com/gephi/gephi/wiki/Fruchterman-Reingold
[47] https://www.tableau.com/

interest that automatically filtered the data and as she kept taking decisions and went down on the visualization. When filtering from the broad top to the specific bottom, would end up with one dataset that by one click could send her directly into the webpage of the newly acquired dataset.

The two visualizations[48], had regarding their differences, slightly different outcomes, but the overall operationalization was the same. Starting with all the data visual, the user can now filter the visualization by clicking on modularity groups, tags, datasets, and so forth.

6 Interpreting the graph

6.1 Municipality Contents

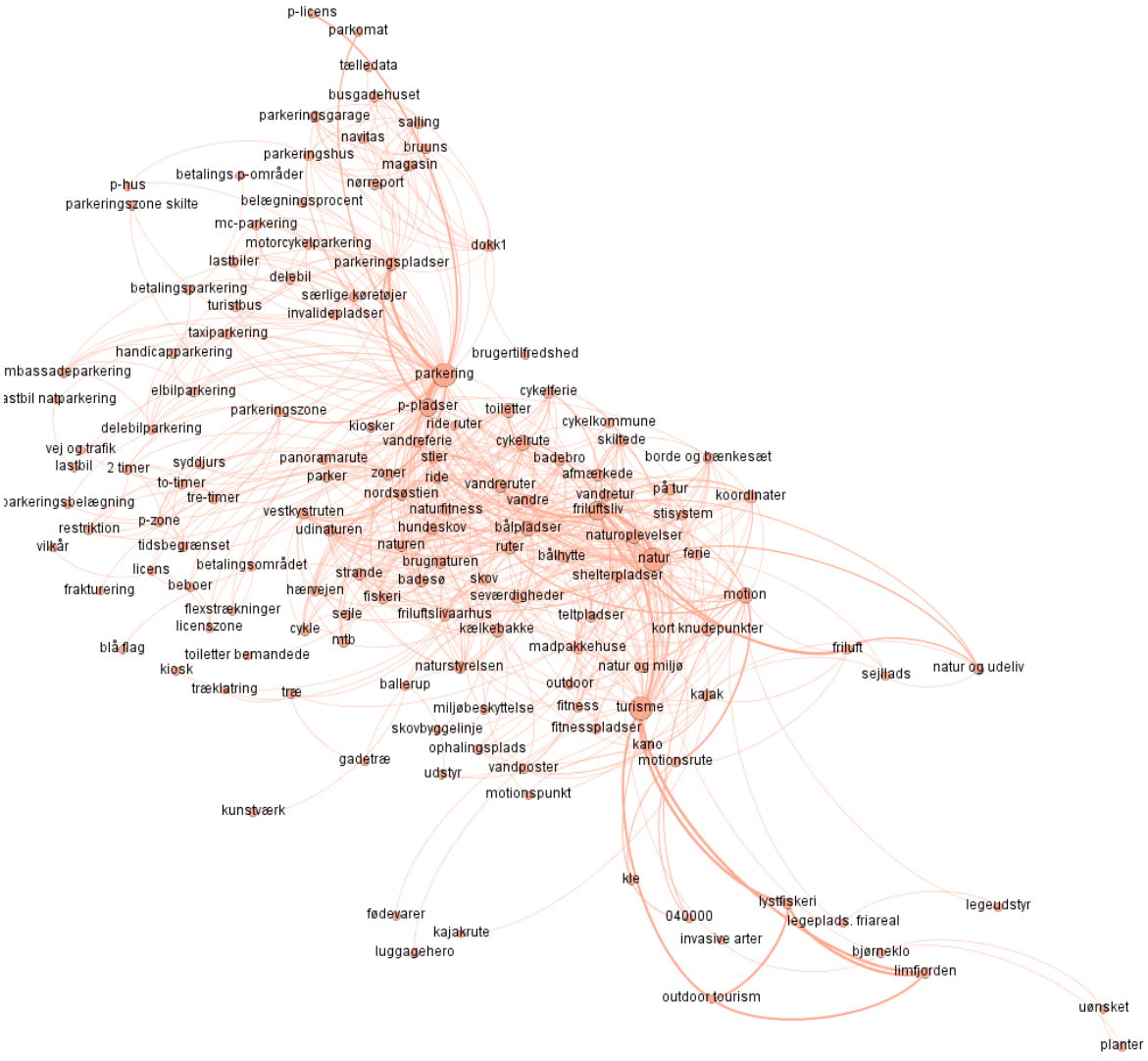
The municipality contents dataset builds on the opendata.dk dataset. I have detected ten major contents clusters that are being used in the municipalities. There are further twenty-two clusters that rather briefly are showed and even further fifty clusters so small that they are not being treated further than this tiny notion. The volume of the clusters is decreasingly getting smaller and smaller as we go along. Starting with the largest in volume and ending with the lowest in volume. I am simultaneously also decreasing the attention offered to the cluster being processed. I am going to present the biggest five cluster[49].

No 1 **-** *parking, nature, and tourism*

The biggest municipality cluster in encapsulated in with the themes, *parking, nature, and tourism,* which also are the three most used tags in the cluster. This cluster no. 1 is a rather big cluster that complicates the process of reducing the cluster and force it into a simplified conceptualization. But the spatialized cluster are often held together by a few bigger nodes that thereby are defining the cluster. This big cluster could have been separated into two clusters where "tourism" together with concepts like *kayak*, *fitness*, *sailing*, *toboggan run* and *outdoor* are forcing the cluster downwards and "parking" in the opposite end with *shared cars, parking zones, roads,* and *traffic* pushing it upwards in the two-dimensional space. The cluster is held together as one big cluster because of the node "nature" that is pulling the cluster together because of

[48] https://public.tableau.com/profile/anders.grundtvig#!/
[49] See Appendix 3 for the more clusters

outdoor/mobilisation dataset like "ud-i-naturen"[50] that contains data both about nature experiences and the infrastructure around it with car parking, bike lanes, and public toilets information. This cluster is therefore defined by the combination of the three aspects of nature experiences as a uniting force. Some of the other big nodes are bike lanes, outdoor life, coordinates, and fishing. A big cluster about *outdoor mobility and activity*.

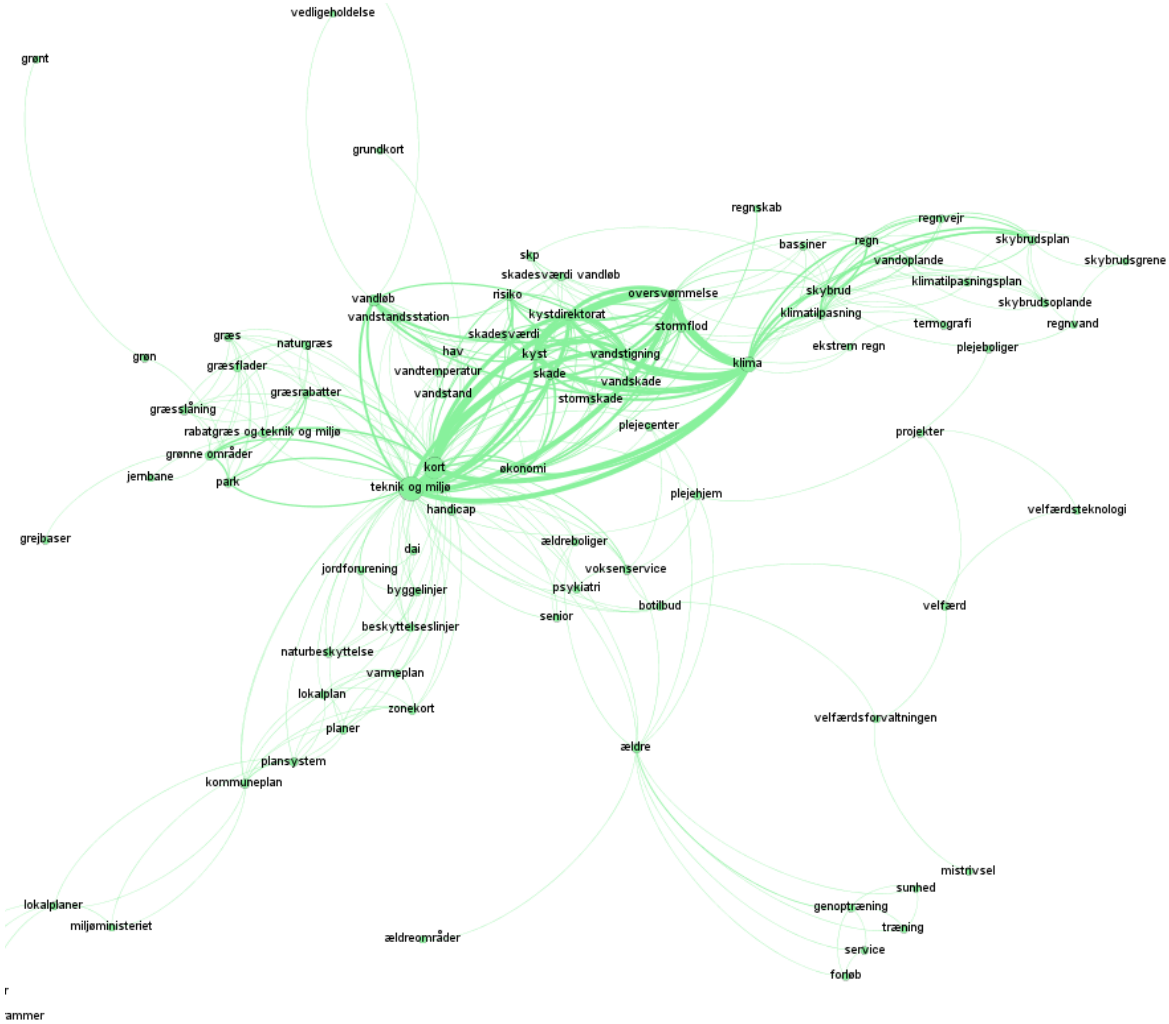


Visualization 4 - opendata.dk no. 1

[50] Translated to out-in-nature

No. 2 - *technology, environment, maps, and climate*

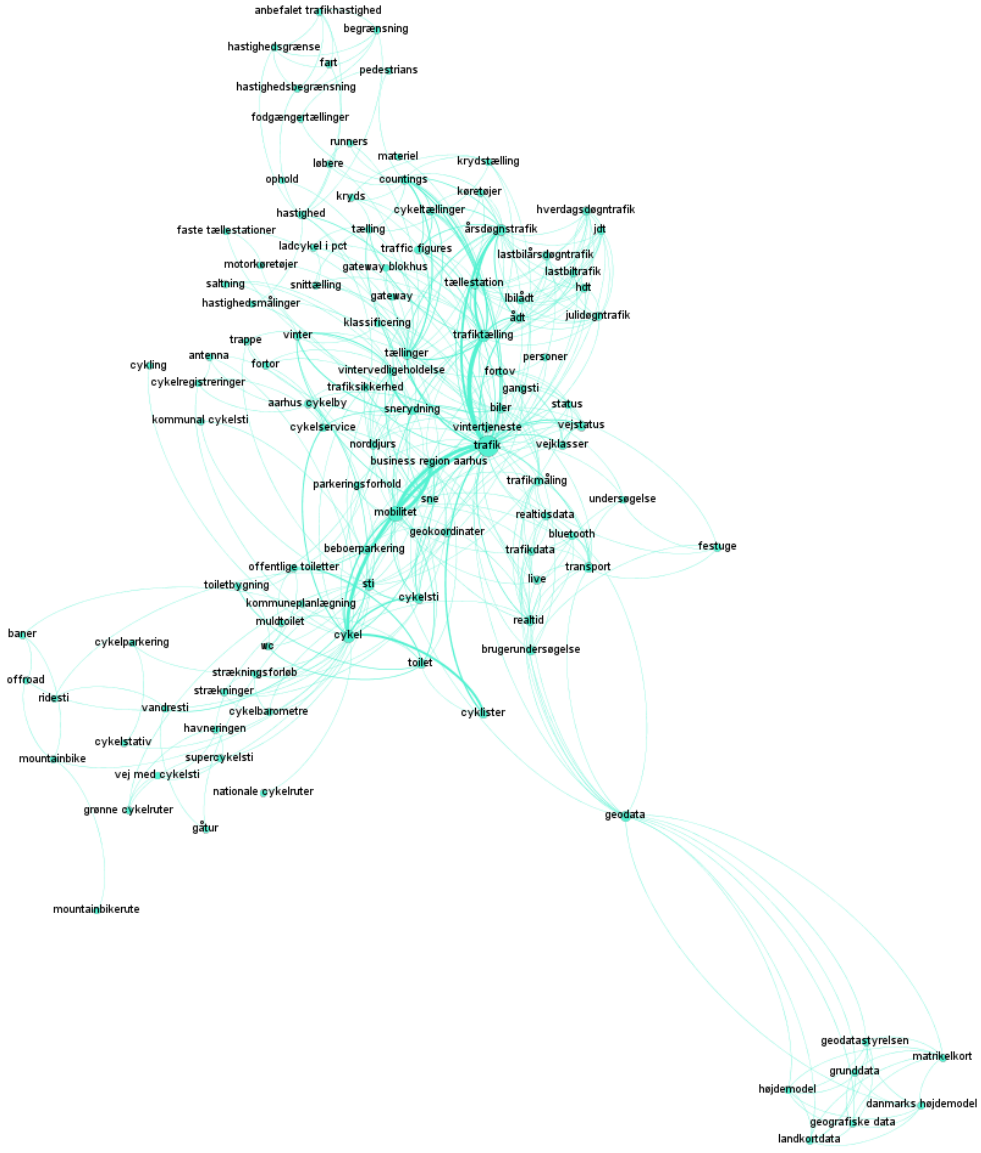
The second biggest cluster in volume concerns *technology, environment, maps, and climate*. This cluster is held together by a common interest in environmental concerns. I did not manually color the cluster, so it is quite incidental that the color for this environmental cluster turned out to be green. Some of the bigger nodes are floating, coastline, ocean, and cloudburst but also smaller sub-cluster appear with nodes connected to the big node, technology, and environments, like a small sub-cluster of elderly concerns of health and services. A sub-cluster about environmental protections that unite the city planners to the cluster. And also a sub-cluster about grass and green areas.



Visualization 5 - opendata.dk no. 2

No 3 - *traffic, mobilization, and bikes*

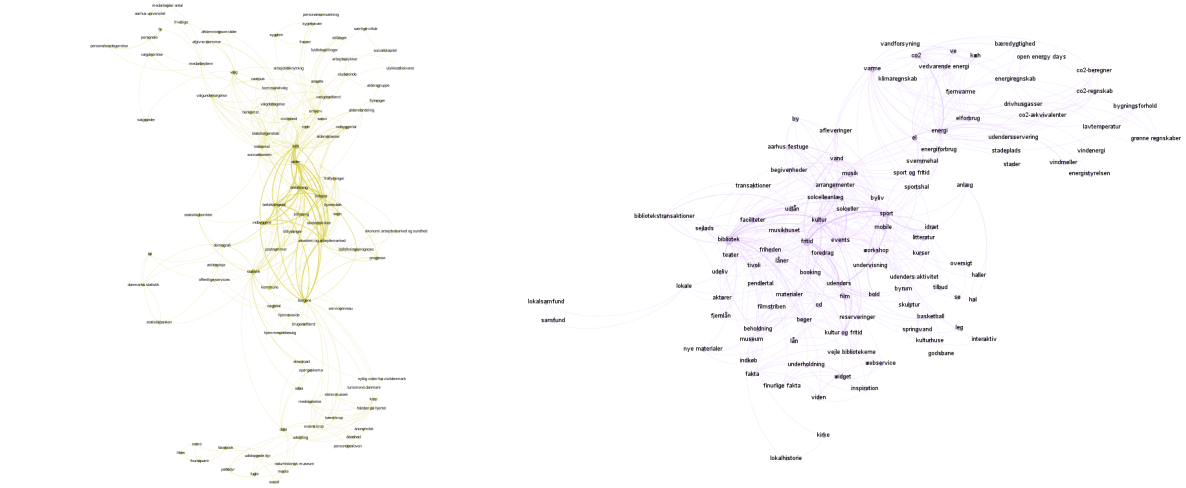
As with the first cluster are we here back to themes of transport and mobilizations just more focused on geodata and traffic counting. The bigger influential nodes are *traffic, mobilization, and bikes.* This cluster is quite united around traffic counting, especially bikes but also other vehicles. There is a sub-cluster, in the bottom, exclusively concerning geodata. And a more integrated sub-cluster connected by *bikes* concerning lanes and routes.



Visualization 6 - opendata.dk no. 3

No. 4 - *citizen, age* and *sex*

The fourth biggest municipality cluster is concerning *citizen, age,* and *sex* with big nodes like citizen numbers, statistic, economy, election, and population. This cluster is the least physical space oriented cluster, and are instead about society related concerns with socio-economics and demography in focus.



Visualization 7 - opendata.dk no. 4 and 5

No. 5 - *spare time, library, culture, and sports*

This cluster is centered around the terms *of spare time, library, culture, and sports.* The cluster is faulty being held together with the sub-cluster in the upper right corner with the term *energy* and *heat* that unite the sports terms with electricity consumption terms. This is one of the issues of this type of text-based topologies data analysis. If the primary purpose for this project was to map clusters of open gov data should these types of misinterpretations had been fixed.

6.2 Ministry Contents

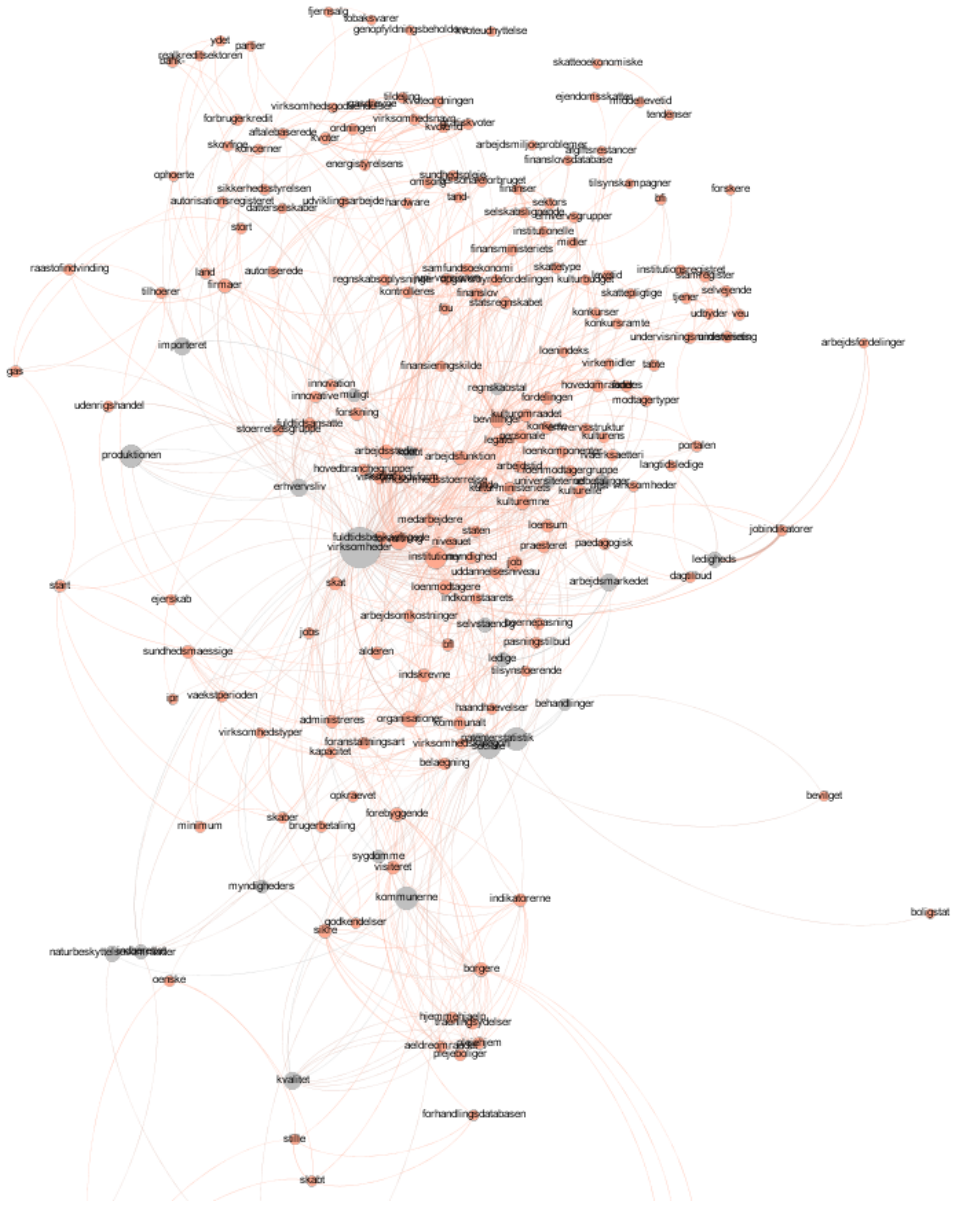
The ministry contents data is built on the rigsrevisionen dataset.

I have again detected ten major contents clusters that here are being used in the ministries. There are further twenty-two clusters that rather briefly are showed and even further fifty clusters. I am here again, going to present the biggest five[51].

[51] See Appendix 4 for the rest

No. 1 - *companies, management, social, and tax*

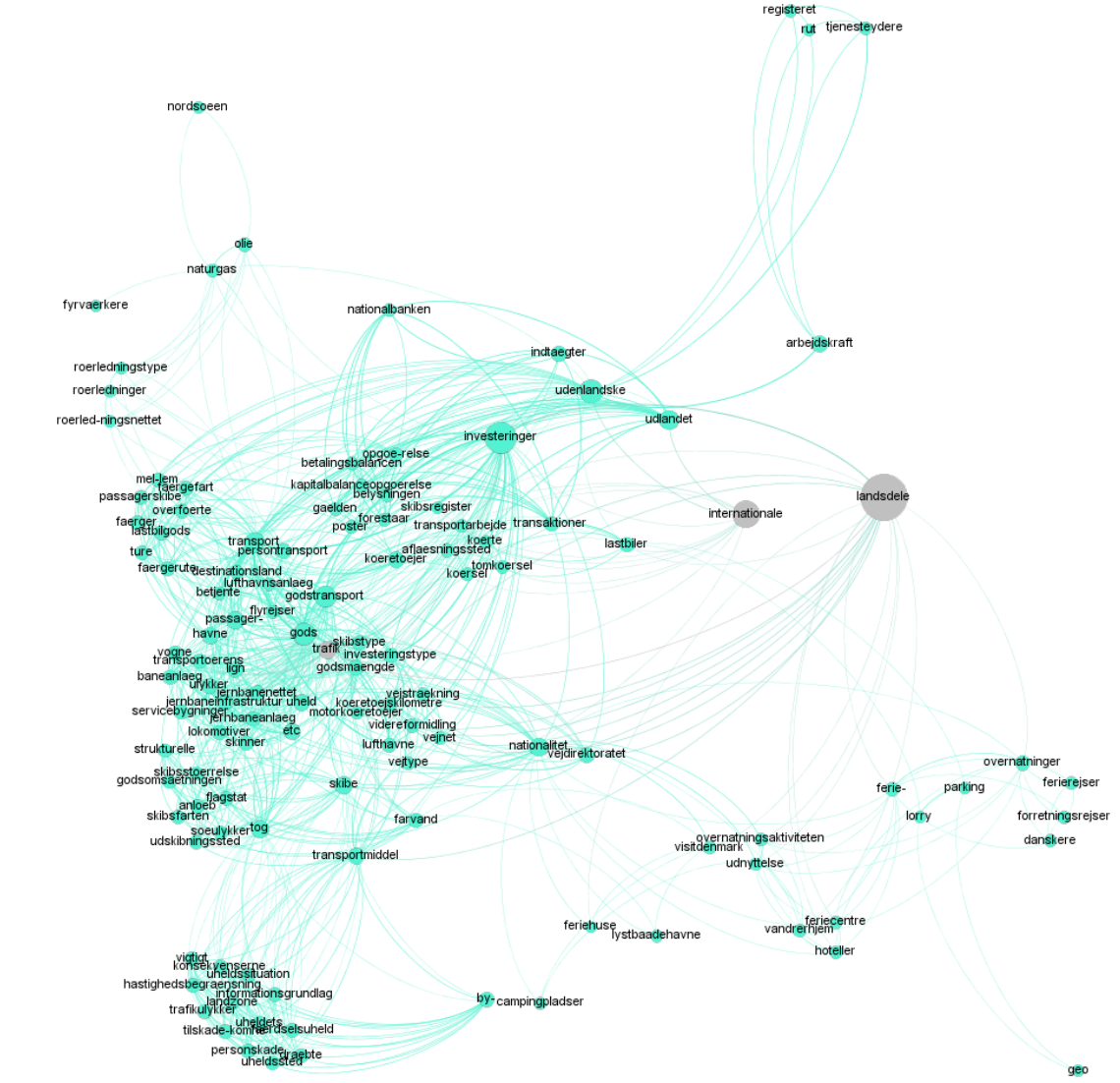
This is the biggest ministry generated a cluster. The cluster is held together by a very broad variety of company/business/finance/trade related themes in the top of the visualization. The bottom of the visualization is mainly about administration and management, and the top and bottom are held together by common concepts of workload, employment and the labor market and the economics and administration in social counseling. Despite the fact that the size of this cluster is the topics represented rather homogeneous.



Visualization 8 - Rigsrevisionen no. 1

No. 2 - *international, invest, airport, and transportation*

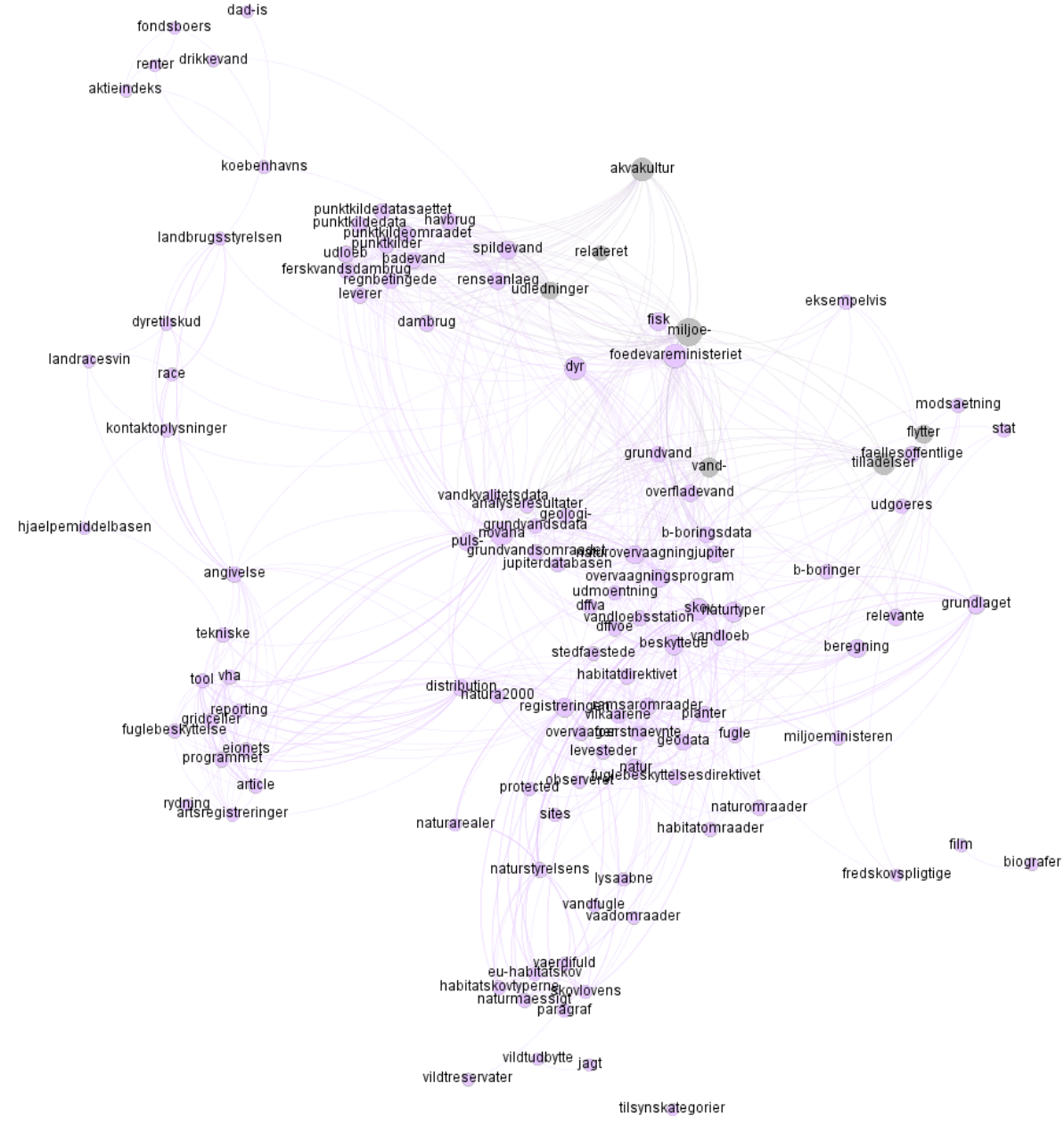
Most of the clusters are highly determined by statistikbanken, but this second biggest cluster is more or less only data from statistikbanken. This cluster is statistics about international finance and transportation and not just as airport transportation, also transportation on trains, ships, roads. The cluster covers the transportation infrastructure. There is one major and a quite dense cluster in the middle and smaller sub-cluster around it. One small sub-cluster in the bottom is about work accidents and speed limitations.



Visualization 9 - Rigsrevisionen no. 2

No. 3 - *fish, forest, nature, mapping, and protections*

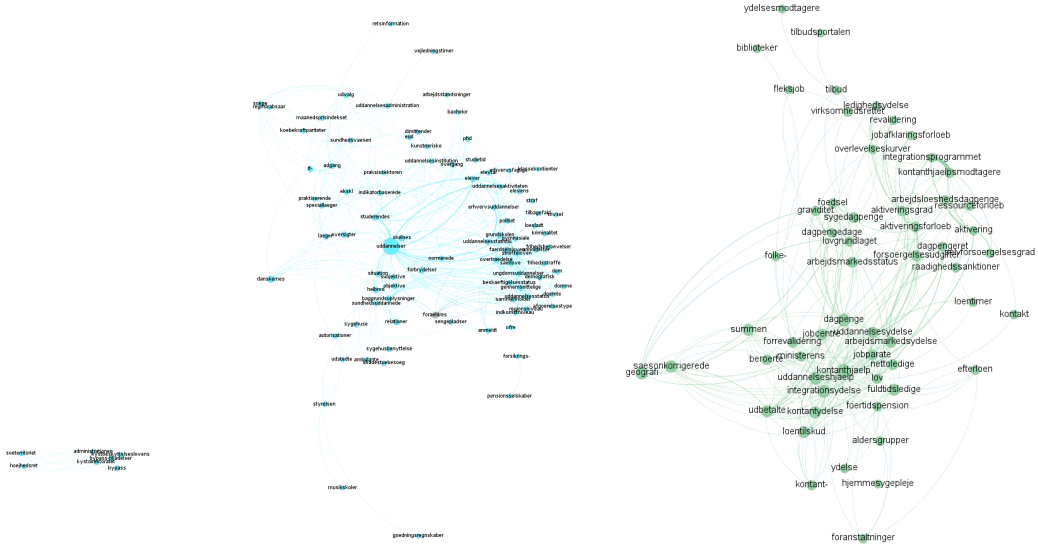
This cluster is actually one of the few clusters that are being mobilized without any major influence from statistikbanken. The cluster about nature-related topics is dealing with nature in the scope of inspections, protections, permissions, surveys, and control. There is minor sub-cluster, e.g., one about water-related nature-phenomenons like swimming and freshwater fish farming. And another sub-cluster about birds protections and bird areas.



Visualization 10 - Rigsrevisionen no. 3

No. 4 - *education, school, student, and criminal*

This cluster that maybe should have been separated into two cluster concerns two major topics one related to educational situations, like education, schools and students and the other major concern is related to statistics of criminal behavior and offense. The two clusters are closely situated in the graph, but a closer examination reveals that the criminal cluster frequently uses terms like education and social conditions in their statistical comparison.



Visualization 11 - Rigsrevisionen no. 4 and 5

No. 5 - *cash benefits, employment activation, and sickness benefits*

This cluster is truly about social help and the social welfare system. The major topics are *cash benefits, employment activation and sickness benefits, and so forth.* The data in this cluster is almost issued by "The Danish Agency for Labour Market and Recruitment" which makes perfect sense when examining the cluster, a quite homogeneous cluster indeed.

6.3 Sum up reflections

What open gov data is out there? This rather simple question that earlier showed to be rather difficult to answer is maybe getting a bit more manageable to gasp with these data visualizations and descriptions of the groupings. These visualizations show connections between datasets to generate clusters that offer *contents* insight about the datasets and hopefully helps to qualify the clarification to the concern of data contents.

7 Platform logics

I am in the following showing why it is important to understand the logic of the platform to deal with the issues related to barriers of open gov data. Depending on the data platform and data portal where the data are being issued and managed the data, especially the metadata, are different and are similarly also structured differently. Data platforms are not just anonymous helpers that provide the technical circumstances for data to be stored and distributed. Data platforms are built, intentionally or unintentionally, with a technical design for data management, data access, and data distribution - this can be acknowledged as the *inherent logic of the data platform*.

I have in this Master Thesis worked with different data platform and datasets. As previously shown, I have focused my attention to the portal on opendata.dk and Rigsrevisionen open gov data project[52]. The opendata.dk portal[53] is built on the open source platform CKAN that is developed to manage open data (Winn, 2013). CKAN is acknowledged as a powerful and mature data management platform in developer networks (Winn, 2013), by scholars adopting the technology (Kučera et al., 2012) and by open source contributors (Costa, 2019).
When contemplating the platform on behalf of the metadata it contains, it begins to be noticed how the different platform is designed with different logics. The platforms ask the data issuer to answer a variety of questions that, afterward, are stored as metadata together with the dataset to communicate and structure the dataset. These questions that the platforms ask differ from the data platform to data platform. And even different CKAN platforms are asking different questions, since the questions that each of the portals are asking have to be coded into the platform by the designer. All CKAN portals do not necessarily contain tags, and the tags can even be tagged in different ways either in freehand or by predefined tags, all chosen by the developer (docs.ckan.org, 2018).
When interviewed datafordeler.dk, we talked about how they have decided to design the data platform on dataforder.dk. They told me about their data distribution strategy and made it clear for me that they had never considered grouping the data with tags vocabularies, as I showed them Opendata.dk did. They explained that they were designing the datafordeler.dk platform to

[52] Rigsrevisionen is not a data platform but a temporary open gov data project offering insights and datasets with statestik and links to concrete ministry issued open gov data.
[53] I here differ between the "platform" (which is CKAN), and "portal." (which is opendata.dk - that is built on top of the platform) "Portal" is thereby the specific opendata.dk portal and "platform" is the broader underlying technology.

accommodate their internal needs and requirements for data storage. A crucial aspect of the platform structure of dataforder.dk was to accommodate open gov data in different stages of sensitive data with an immense focus on security built into the platform. The data on datafordeler.dk is therefore secured behind appropriate levels of user permissions.[54] This illustrates that data platforms are built and developed to accommodate specific needs and obligations and thereby are building different inherent logics into the platforms. These logics further suggests that the data and the metadata on the different platforms are more or less incommensurable with each other and thereby makes it difficult to unite all open gov dataset in one big comprehensive data platform as the agency for digitalization is working on[55].

In order to facilitate an emerging and continuous use of open gov data, the users have to become familiar with the data. But with 84 different open gov data platform[56] where most of them are built with their data logic, it is quite difficult for the users to comprehend and use the data across the platforms. Especially unstructured data platform without any API data access is according to Jakob and Alex pointless to work with.

"Who is the intended user of all these data platforms?" Jakob asks me. "The Partnership would argue that you are" I reply. Jacob continuous, "But then I do not understand why they make it so difficult for me to work with, with all these different platforms, accesses and structures. I really do not feel like the intended user of this product"(Author, fieldnote, Sp8ces, 2019)

The frustrating feeling of not feeling like the intended users of open go data is a theme that many participants have raised and something we are going to scrutinize later on, but for now, it is interesting to look at this frustration from a data access point of view. Jacob has been communicating with different database API's during his short career, but this communication, he explains, is difficult and takes a lot of time to manage. Lars is also making this point of time and energy consuming when managing and communicating with different data sources. The data adaptation of data platforms and the act of using them frequently depends on the aspect of becoming familiar with the data and the platform.
The logic of the platform also facilitates the way that the data is supposed to be looked for and found by the users.

[54] https://datafordeler.dk/vejledning/om-datafordeleren/moderne-infrastruktur/
[55] https://arkitektur.digst.dk/node/548
[56], This number is from the rigsrevisionen dataset and is actually a minimum of platforms where, e.g., opendata.dk is not included.

8 Information Search

Open gov data is with multiple accesses, structures, and reasons for being developed, a difficult technology to encapsulate as one united field of interest. When users (whoever they are, more on that later) are searching for open gov data, the data search is either being applied through specific open gov data platforms known in advance or at widespread search engines like Google Search. The whole concept of open gov data is built on massive amounts of different forms of data, distributed across platforms to be interpreted externally of the government. The success of this type of concept relies on well-designed interfaces for data access and entry points.

8.1 Data encounter and data search

Finding relevant open gov data has many similarities to finding interesting material at the physical public library; *you get in, looks around, takes different material down from the shelf, looks at it, and if it works - you bring it back home*.

Lennart Björneborn (2008) is investigating how users of a public library are accessing materials throughout multiple entry points. He is theoretically dealing with two main concepts for accessing the data, goal-directed, *convergent* information behavior and explorative, *divergent* information behavior. Convergent information behavior is the act of knowing exactly what you are looking for, just using the effort actually to find the right material. Divergent information behavior, on the other hand, refers to the act of not knowing what you are looking for beforehand, making a search-journey into unknown fields of *terra incognita* (Stark, 2008). Björneborn (2008) is highly interested in the user behavior and investigates situations where the library user is leaving the library, equipped with unintended material that during the visit called for interest and attention, a phenomenon he categorizes as *serendipity.* He is presenting three affordances for the library to foster and inspire the presence of serendipity, *diversifiability* concern how diverse the content and access are. *Traversability* concerns the access and how the access point is inviting for usage, *sensoriability* concerns how the senses are being stimulated. These approaches are later going to be the foundation for discussions when compared with concrete open gov data accessibility, but for now, is it enough to know about the existence of them.

Stark (2008) is contributing to an innovative dimension in information searching, where he argues that *challenge searching* is related to innovation. *"The most important component of innovation is a process that is not directed toward the solution of well-defined problems."* (Stark, 2008).

Stark (2008) is dealing with information searching and is investigating how obtaining and dealing with information is affecting knowledge production and innovation. Stark, inspired by Dewey, is showing how innovative problem-solving are perplexing and challenging situations that cannot be solved with what he acknowledges as, *simple searches.* He differs between simple searches that solve well-defined problems and *challenging situations* that require the entrepreneur to navigate the unknown field of uncertainty; *terra incognita* (Stark, 2008). *"[…] faced the challenge of knowing that sometimes you must search even when you don't know what you're looking for."* Stark (2008) argues that the innovative strength of exploration is the ability to recognize new combinations of concepts that not yet are accepted as a concept but are emerging in the process of the open-ended unfamiliar search.

Björneborns (2008) *serendipity* differs from Starks (2008) *exploration* in aspects of finding material where serendipity is aiming to find something you didn't know you were looking for, exploration aims to complexifying and creating new open-ended situations.

Dewey argues that it is artificial to think that perplex situations are co-occurring with defined problems and situations. Instead, problems are here seen as the "solution" of the perplexing situations where Dewey argues that reflective thinking is the key to findings (Stark, 2008). Dewey is grasping the concept of problems, never as problems but either as *assigned tasks* or *perplexing situations* are interesting in situations of recognizing different kinds of searching. Innovative problem solving relies on the ability to search beyond concepts of assigned tasks and well-defined problems and instead should foster the search ability to investigate perplexing situations and to explore unknown territory in the goal of exploring new concepts.

Related to this Master Thesis, I am especially interested in the theoretical search terms and concepts presented above concerning conceptualize and discuss open gov data barriers according to the access and entry point as searchable data. In my interpretation of Stark (2008) he argues that if open gov data is intended to solve perplexing situations, it is incomplete to structure the data behind simple search functions. Björneborn (2008) argues on a practical level that open gov data should be analyzed in the scope of having a divergent and a convergent

interpretation where he is offering hands-on examples of supporting the presence of serendipity that he would argue is an important aspect in information searching

Open gov data access can in various aspects be perceived as to accomplish the same territory of success as libraries have been demonstrating and refined for ages, *the capacity to communicate and distribute materials.* We are later using the comparison with the library as a data entry example. We are in the scope of open gov data going to look at the material communication and distribution taking place in the light of drawing inspirational lines between the two entities.

9 Serendipity, as the main object of data encounter?

An intriguing part of open gov data is the access entry of data encounter; where and how does the user find the open gov data? There is a variety of different strategies to arrange how the user's data encounter is taking place in the landscape of open gov data. One of the most recycled strategies for open gov data initiatives is to advertise possible access points for data encounter by focusing on user cases[57][58][59] an approach that could be acknowledged as *the single case entry point*.

These single case entry points are an attempt to accommodate the user's uncertainty of the usability of open gov data, by showing that other companies and initiatives are using the same data, and therefore it must be interesting. This data encounter strategy leaves the current and potential users with a (hopefully) enlarged appetite for open gov data but leaves them without any concrete ideas of where and how to find and encounter the data in more general terms. I have located 84 different data platform for data encounter with different accessibility and structures. I here provide a few examples of data entry strategies on the different platforms; "Datafordeler.dk" requires a login and have a special designed API[60]. "Nyidanmark.dk" only provides non-structured string-data (normal text) that can be eyeballed on the page[61]. In order to pull data from "dk.vandmodel.dk" the user needs to ask for the permission on email, and the access is given "ad hoc" as they write[62]. Statistikbanken.dk data can be accessed directly on the platform with filters and searches, and also through a special designed API[63]. Retsinformation.dk requires two registration and approvals; a company certificate (VOCES) and a web service method authorization[64]. Kortforsygningen.dk requires a login, and if you want custom made data, you have to request it and wait for up to two-three days. Comprehensive data pull obtained through a FTP-server otherwise eyeballed at the browser[65]. And so forth with different data entry approaches on the different platforms.

[57] https://challenges.dk/da/urbantechchallenge/aabneoffentligedata
[58] https://data.virk.dk/inspiration/cases-0
[59] https://www.kl.dk/media/10948/34-laes-pjecen-om-de-smaa-aabne-datasucceser.pdf
[60] https://datafordeler.dk/vejledning/brugeradgang/brugeroprettelse/
[61] https://www.nyidanmark.dk/da/Om-nyidanmark/Om-internetportalen
[62] http://dk.vandmodel.dk/adgang/
[63] https://www.dst.dk/da/Statistik/statistikbanken/api
[64] https://www.retsinformation.dk/offentlig/vejledninger/Retsinformation%20web%20service%20vejledning%20v3.pdf
[65] https://kortforsyningen.dk/indhold/download-af-data

Data access is in this perspective beclouded by uncertainties of how to get the data and where to find the data, some with API's some without. The focus in APIs is further interesting because of acknowledged as an important element in the digital infrastructure of the city (Raetzsch and colleagues, 2019).
This uncertainty was by different participants at UrbanTech Challenge dealt with in different ways. The most common way of overcoming this issue was by asking "the help deck of the internet" for consultation; the search engine "Google Search." Another popular approach was to look for data that the participant knew about beforehand. Some participants ended up using data that they had requested for the data owners, but mostly the participants ended up using the first dataset presented for them that they somehow could relate to, either through Google Search, the data owners or previous knowledge. Roughly speaking, *simple search*, data are known beforehand and requested data, can be acknowledged as a data exploration strategy that heavily relies on pre-conceptualized ideas and could be categorized as *convergent* data encounters without explorations elements.[66]

The *terra incognita* (Stark, 2008) suggests that the most crucial point of innovations is the process of moving in new unknown territory and argues that simple solutions to well-defined-problems are not fostering innovation. This taken into mind, it becomes interesting to explore alternative data encounters than solely the convergent-driven data search. Björneborn (2008 and 2017) recommends the development of an explicit information encounter design, focusing on combining convergent and divergent approaches, namely serendipity.
Serendipity is the concept where you arrive with one key idea or goal and leave the place with another and better idea. Serendipity affordances at the library create an innovative design space that inspires the users in new directions to make new and better decisions. In this scope, the library is not just aimlessly shelves of books as a physical data storage platform that easily could be replaced with e-books, but rather an influential inspiration space with opinions and recommendations, designed to open new exploratory paths for the users. Björneborn (2017) is, with his three accordances that facilitate serendipity recommending how serendipity can be achieved and managed.

[66] Reality is of course not that simplistic, simple search eg. is only convergent when searching for specific narrow terms like "adresse data", but when the search query is broader with open concepts like, "åbne offentlige data" then the convergent/divergent division is difficult to outline.

The concept of encountering open gov data on a platform share similarities with encountering materials on a library. Both the library and the open gov data platform are storages for materials that are supposed to be located and found by the users. But libraries are designed to enhance the communicative role for the materials and create new aspects of encounters. The portal on opendata.dk is an example of a data storage design focusing on the convergent data encounter with filters and a search bar.
In the following, I am going to elaborate on Björneborns (2017) three affordances for serendipity in relation to open gov data and is thereby scrutinizing how open gov data encounters can get inspiration from libraries and is investigating what happens when open gov data users are exposed for a serendipity inspired data encounter.

Björneborn (2017) argues that serendipity *per se* cannot be achieved or designed, but have to be obtained by designing the building blocks, the serendipity affordances. Björneborn's (2017) distinction between the *unplanned* and the *unexpected* is in this case relevant. The participants are participating in the workshop *expecting* to be introduced for new data, but they did not *plan* to find the precise datasets, the serendipity taking place here is unplanned but not unexpected.

9.1 Serendipity at the workshop

I have in the preparations for the workshops, by designing the visualizations and the space for the visualization to be interpreted, attempted to direct the workshop design with the inspiration of the serendipity affordances. One of the most important aspects of the workshops was to facilitate discussions about what kind of data that are relevant and interesting for the participants to work with. In the definition of serendipity, Björneborn (2017) argues that the three key terms are *unplanned, encounter, and interesting.* The three terms were similarly interest aspects for my workshops, to experiment with a data *encounter* that fostered the participants to *encounter unplanned* and *interesting* materials.

We are again situated in TANT-lab. We are still in the starting phase of the interview, and we are discussing how Lars made his first data encounter during UrbanTech Challenge. I am reminding Lars about some details of his winning concept at UrbanTech Challenge by reading out loud from Lars' UrbanTech Challenge hand-in. The hand-in involves the data sources that Lars proposed to use and re-combine. Lars has earlier studied and work on urban design projects where he also got his knowledge about open gov data sources from. When developing his concept, Lars was

browsing well-known data sites, like 'kortforsyningen,' for concrete data he knew was out there. In the process of browsing the web, Lars ended up at opendata.dk where he searched for a potential dataset for his concept. He came across a dataset about trees in Copenhagen that he considered using somehow, but unsure how this dataset could help his concept. He continued with his first idea about an oblique photo from 'kortforsygningen.' "The thing about open gov data in the data concepts that I have been working with," Lars tells me, "is that I know that I can collect a lot of different data, refine them and combine them in different ways that I am pretty sure would create some interest. But things tend to become more and more complicated when you are dealing with a lot of different kind of data and data sources. I would rather prefer to find a few, but really interesting data that I could work with that would make my product easier to work with both for my clients and me."[67]

We are here introduced for the most common and straightforward way of finding and looking for open gov data; knowing the theme of what you are looking for, browse a bit around the different platform, and hopefully find the desired data.
Lars is here searching for datasets on opendata.dk but decided to continue working on his concept with the oblique photo from 'kortforsygningen'[68]. He actually *unplanned* encounter a dataset about trees of which he did not know existed - but he also did not find it that interesting either.
Let us fast forward to the part of the interview with Lars, where we again are interpreting data from opendata.dk but this time mediated through data visualizations in the workshop setting at TANT-lab.

After spending a longer period explaining the dataset from the basics of the visualizations and after demonstrating how the visualizations were produced, we are systematically going through each of the opendata.dk graph cluster visualizations. Lars is showing a lot of interest to the cluster visualization group no. 6, containing, 'buildings,' '3d', 'citymodels' and 'areas,' so we end up spending a lot of time with this visualization on the big screen. "So what data do we have in this cluster?" Lars wonderingly asks, and continuous themselves, "A sub-cluster with '3d city model' that is interesting, why didn't I come across that when I looked myself? Have I just searched for the wrong terms? I think this is certainly useful. 3d city models, buildings, SketchUp, urban planning,

[67] Author fieldnotes, Lars Hartmann group, 2019
[68] https://skraafoto.kortforsyningen.dk/oblivisionjsoff/index.aspx?project=Denmark&lon=10.2027929&lat=56.1277927

that all might be data I can use." Lars ends up replying his own wondering question. "This is amazing; I really want to dive into some of this data and try to see if I can make some sort of product on behalf of this data. There seems to be so much! If this had been the starting point for UrbanTech Challenge, there would have been much better solutions. It is for sure."
The role of being the one making notes have now changed. Lars is making a lot of annotations in his small white book. We are seated facing each other with a small recorder in between us, snitching our conversation. Lars, not intimidated by the recorder, asks me to wait for a bit, while he finishes up typing. Anticipating Lars' next reaction, I am pleased to experience that the visualizations help him encounter new data that he finds interesting. Lars kills the silence "OK. I am ready again" he says, while still typing his book. "It is really interesting that there is data about SketchUp, it is a Google 3d data format that I have been working with before, it would be cool if we could find some data in that format.". I am, On the big screen, opening the opendata.dk portal and search for SketchUp. We quickly realize on the portal that the SketchUp file format is '.skp,' and with this knowledge, we can now filter the entire portal exclusively for .skp formats. We find some data sets about 'Aarhus island.' Lars is enthusiastic. "Okay! So they have made the whole of Aarhus island as SketchUp files. It is exciting. I just have to get all of that data and check it out."[69]

Lars arrived at the workshop equipped with a predefined and definite vision of the open gov data he intended to use in his concept. He knew that he could transform the oblique photo from 'kortforsyningen' into urban 3d models that he then could refine and combine with other well-known data. With the oblique photo concept Lars managed to win the secondary prize in UrbanTech Challenge, so, not a poor concept after all. But an interesting situation at the workshop with opendata.dk graph visualization no. 6[70], was to experience how Lars almost felt a release when introduced to complementary and relational data to his concept. Now, he had extended flexibility to shape his concept not limited by the determination of the availability of data on kortforsyningen and not limited by technical competences of transforming the oblique photo into 3d data.
The data journey of finding interesting data is, here at the workshop, approached from a serendipitous direction where the participant is encouraged not to *search* for planned data but instead encouraged to *encounter unplanned* data on the visualizations. This serendipitous journey includes convergent search aspects, where Lars, qua his background, was looking for data, related to urban design concepts. This led the attention to slide no. 6, where he is a

[69] Author fieldnote, Lars Hartmann group, 2019
[70] Appendix 3

divergent encounter, was confronted with a variety of different tags in different sub-cluster all calling for his attention. The name of the clusters was pointing and navigating the data visualization by highlighting the most used tags. After being pointed in a direction, Lars had to explore the visualization by visually interpreting to obtain further information about the data available.

The above-outlined serendipity visualization experiment does insinuate that designing open gov data encounters for serendipity affordance enhance the user's ability to locate interesting data. But serendipity is also recognized to affect different people differently (Foster and Fort, 2003) and this experimental study with serendipity factors is also a rather small study and therefore difficult make a profound conclusion on behalf of. Björneborn (2008) is further highlighting that designing the facilitation for serendipity factors and affordance is not only about enhancing the factors but, importantly, also not to counteract the factors. The *sensoriability* affordance is, e.g., empathizing, that the designer facilitates and empower the senses to play an active role in the encounter (Björneborn, 2008). The visualizations employed in the workshop are solely digital and does, therefore not activates any senses other than the commonly stimulated, *visual* sense. It could be quite interesting to further experiment with the sensoriability affordance in open gov data projects. An important distinction in data visualization is between the *exploratory,* the *confirmatory,* and the *explanatory*. The data visualizations in this Master Thesis have been focusing on the visualization to facilitate explorations and discoverability elements. Since there is no correct way of exploring the data visualization, and since there is no uniquely message to be communicated, the data visualizations in this Master Thesis are encouraging an exploratory data analysis.

We have here explored serendipity affordances during the workshop and explored how inspiration form serendipity can stimulate users of open gov data to encounter and discover new unplanned data. We are in the following going to explore how serendipity affordances are present in open gov data encounter on a broader scale than in the workshops, based on the portal at opendata.dk where we are exploring how the portal in-itself is stimulating serendipity affordances and are connected to the surrounding infrastructure of promoting open gov data. The following broader serendipity exploration is related to scoping serendipity and especially *counteract* for serendipity as socio-technical adaptation barriers for open gov data initiatives.

9.2 Open gov data and serendipity - *or the lack of the same*

Open gov data databases like the portal on opendata.dk are being framed and are framing themselves as data storage projects. Eg. opendata.dk description of the portal, *"The initiative aims to gather and exhibit freely released datasets from all over Denmark on one nationwide portal."* [71] The term *exhibit*[72] is the interesting term to scrutinize in this case. The question, at least for me, is how to persuade the underlying substance of this term. Should open gov data platforms, be platforms for data to be uploaded on and succeeds by being a warehouse for data uploads? Or should the platforms play a more active communicative role and also be evaluated on their success of communicating the data to the users? It comes down to the distinctions between, *storing, displaying, or curated exhibited* the data on the data portals.

During the facilitation of UrbanTech Challenge when we attempted to motivate the participants to explore open gov data, e.g., on opendata.dk, it became noticeable that the participants accessed the data portal with, in general, two different encounter attitudes. Either the participants opened the portal and with no preconceptions was encountering the portal, or the participants were encouraged to search for data on the portal by inspiration sources elsewhere. In the former approach, the participant was motivated and encouraged to encounter data on the portal by the design of the portal itself, *internally*, in the latter approach, the participants found this motivation elsewhere *externally*. In the internal example, the portal is holding the responsibility to facilitate inspiration and navigate the users to explore interesting data as a curated data exhibition. In the external example, the primary role of the portal is to support an organized structure that allows the participant to find the desired data. We are here presented for two different data entry approaches, both taking place on opendata.dk but with two different data entries expectations for the portal.

Mathiassen and Lindstrøm (2015) declare that the portal on opendata.dk, not just is a data storage portal but also a portal that provides a comprehensive overview of open gov data among the municipalities. The notion on *overview* is interesting since it indicates that the portal on opendata.dk also is supposed to accommodate data access motivations internally.

[71] http://www.opendata.dk/ Accessed 26.05.2019
[72] Author translation from Danish "udstille."

The focus on internally and externally motivation factors are in relation to acts and counteracts of Björneborn's (2017) serendipity affordances, important starting points to keep in mind. When we are evaluating serendipity affordances in open gov data, we thereby have to differentiate between the scope of scrutinizing the portals as an actor situated in a larger infrastructure of actors, or as an infrastructure in-itself, independent of the surroundings. The situation earlier with Lars eg., showed how he ended up using a combination of different data entries in his concept. Lars pre-conceptions and data experience were used as an external data entry. I introduced him for data visualization that similarly was an external data entry. We then used both his pre-conceptions, data experience, and the data visualization on the portal to find interesting data. In this case, the portal was functional, exactly because of external data entry. When Lars earlier tried to encounter the portal, he did not find any interesting data. This suggests that the internal data encounter does not encourage the user to find interesting data on the portal - *independent of the surrounding*.

The situation is, of course, not that simplistic. The point is just that open gov data serendipity affordances can exist independent of the portal. But no matter where the serendipity affordances are taking place, it is important that the infrastructure is not counteracting serendipity affordances.

The data portal on opendata.dk is built on the open source platform CKAN. The Open Knowledge Foundation[73] that is the original developers and maintainers of the CKAN platform[74] is repeatedly emphasizing that the data platform is a diverse system that accommodates data uploads, storage, management, solr search engine, filtering, grouping, previewing, analyzing, visualizing and further is connected to a strong network of developers that continuously build new accessories and features to the platform[75]. The platform is, despite its diverse systems, more or less built to facilitate convergent data searches.
During the UrbanTech Challenge, I experienced that the opendata.dk portal did encourage the user to click on the filters and anticipate new datasets to appear. The portal also provided the users with meta insights with visual overviews about the most frequently used tags, most active municipality, and a number of datasets.

[73] https://okfn.org/
[74] https://okfn.org/projects/ckan/
[75] https://ckan.org/2018/09/20/open-letter-response/

If accepting the importance of Björneborn's (2017) affordance for serendipity as key elements in information searching and the availability of the materials, the data portal should enhance the divergent encountering features and in doing so, explicitly consider how to stimulate the affordance for serendipity. Let us look at the material encounter experience at Rentemestervej library and let us compare it with the data access on the data portal at opendata.dk.

Entering the library numerous of things fighting for my attention. A variety of books looked at me all at the same time but from different places in this big open room. art constellations hanging and standing, colorful magazines, messy children, and librarians were all in a surprising calm vivar doing their own business.[76]

At the entrance of the library, we experienced a fight for our attention with a lively atmosphere and a broad variety of materials actively positioned to capture our attention. If the analogy to the library is the opendata.dk portal, the library entrance is here the front page of the portal. On the front page of the portal, our attention is, as just remarked above, drawn in different visual direction - encouraged to click on a filter.

I had a concrete seat in mind for my contemplation but decided to take a small detour through the library to get there. Overwhelmed by the calmness of the books, I automatically slow down my movements. There was literally books everywhere.[77]

On the library, we are filtering the materials by physically moving into another area of the library. We can e.g., filter for "books" and "biographies" by moving to that area of the library. On the portal we are simply clicking a "municipality" or a "tags" to filter the datasets, the platform then automatically corresponds. In the moving (filter) process at the library, we are exposed to encounter unplanned material along our way of moving, walking by categories we did not know existed stimulated with *diversifiability* affordances. At the portal, the filtering process, at worse, take up a few second letting us look away from the computer. In the library, we are standing among biographies books. An armchair is inviting us to take a seat while staring at the books.

Here, surrounded by books, another reflection about books stroke me. It is not just the library that communicates the books, the books themselves if given the change, are handling this

[76] Author fieldnote, Rentemestervej library, 2019
[77] Author fieldnote, Rentemestervej library, 2019

commutative role as well. books are designed with outer interpretations: A catchy front cover, that communicates the content of the book, an insightful back cover, that communicates the book with a summary and reviews, and an informative spine, that in a brief and restrained fashion informs the name of the book and the name of the author.[78]

At the library, particular books are exposed with the *catchy front cover* pointing at us, visually communicating its content. The rest of the books are showing their *informative spine* mixed with the masses of books as a big colorful painting, all stimulating the *traversability* affordances. At the portal, after filtering for "Copenhagen municipality" and ".csv" files, 144 datasets are asking to be scrolled through. No particular datasets stand out of the plain page. No dataset is calling for extra attention, e.g., with a *catchy front cover* but is instead all showed with something mostly comparable with a colorless *informative spine* as an anonymous books shelf of ancient books. We are on the portal empowered with the possibility to rearrange the datasets according to different criteria like alphabetical, last updated, and popularity. But the rearranging mechanics are controlled by a small, easy to overlook, button. On the library, the bookshelves are similarity showing a few popular books and a few new books as well, and the library is not afraid of promoting and advertising for their special highlighted books. The presented books are asking to be taken down the shelf, felt, weighted, sensed by the users, scrolled through and The *insightful back cover* to be skimmed, stimulating the *sensoriability* affordance. On the portal, the data cannot be sensed by other means than eyeballed, and the data communication is designed to facilitate nonpartisan systems with the datasets organized in straight, organized lines.

This library was indeed not a simple material warehouse or book storage space; this place was obviously designed for liveliness, not for storing. But it did indeed as well as store materials as just described. Books, magazine, and catalogs were consciously placed on shelves and stands in orderly organized systems. Note the plural "s" on systems. Because in here, multiple bibliometric systems are build and orchestrated together. An alphabetic system, a numeric system, thematic systems, index systems, all are being orchestrated by the design of the library as the main conductor of this material-orchestra. But the materials here are not "just" organized in bibliometric machine logic systems. The first shelf met when entering the library was dedicated to books that other users earlier had read. On the shelf in the middle of the big room are books that

[78] Author fieldnote, Rentemestervej library, 2019

the librarian recommends. And even when sweeping around the bibliometric organized shelves stuffed with machine-logic, randomly selected books are presented with the front cover instead of the more common space optimized book spine storing fashion.[79]

On the library, we experience a considered and designed material-mess situated in structured bibliometrics systems again stimulating the *diversifiability* affordance. The users of the library are visiting the library of different reasons and motivations, some are targeting concrete materials, but others are commuting because of specific events, reading rooms, children amusement, on Rentemestervej library there is in-house citizen service, and normally they also facilitate fabrication and art workshops and studios. On the portal, the users are properly visiting because of increasing open gov data curiosity and target data search.

In this experience comparison between Rentemestervej Library and the portal at opendata.dk, we are presented for the idea that the portal in comparison with the library is favoring *searches* over *encounters* in the data communication and data access. The difference here between *search* and *encounter* is referring to the discussion about the *serendipitous* unplanned data encounter versus the targeted *simple search.* The library has built a space where serval serendipity elements are welcomed and designed for (Appendix 3 and, Björneborn, 2008). The portal on the other hand, despite the CKAN ambition of being a diverse usability platform that should cause it to be more than just a storage (kilde) and despite the overview ambition (Mathiassen and Lindstrøm, 2015), the portal seems to be accessed and used as a portal mostly for simple data search.

9.3 Subset - serendipity and open gov data

Let us take a look at the three affordances for serendipity, *diversifiability, traversability and sensoriability,* to locate related open gov data barriers both specific at the portal but also in general and locate how this is related to the absence or counteract of serendipity.

Diversifiability is described as, *"[..] how does this environment allow a diversity of contents"* (Björneborn, 2017). The interesting and successful aspects of the portal on opendata.dk, and the CKAN platform, in general, is the possibility to accommodate diverse data sources and data contents. As earlier outlined, the CKAN platform is allowing multiple data owners and multiple

[79] Author fieldnote, Rentemestervej library, 2019

data sources, formats, and versions to be stored on the same platform. This diversity aspect was one of the key motivators for the participants in UrbanTech Challenge, and for me in this project, to engage and use this platform in favor of others or none. In more general terms, open gov data is encapsulated in the existence justification of being a diverse data source. This data diversity seems both to animate user interest but also creates data management challenges when trying to comprehend the scope of this data source that also disorients the users in relation to the *concerns of data contents*.
Users of open gov data are optimally being introduced for a diverse data amount and content, but the users are not necessarily encountered with this diversity. I would argue that the environment of open gov data allows a diversity of contents, but the contents are not visible for the users. Björneborn (2017) elaborates on *diversifiability* with an example of serendipity at the country roadside, where the driver of a car unplanned is encountering the environment. If this country roadside is translated to the case of open gov data, the car driver (open gov data users) is passing a diverse landscape (the different data sources and datasets) but is not encountered with the landscape since it is not visible from the car. The driver has to leave the car and enter the landscape and search for specific elements to appear.

*Traversability***,** is described as; *How rich affordances are there thus to move through a given environment and reach different resources?* (Björneborn, 2017). The open gov data environment might succeed in the diversity of the data and data sources but struggles to communicate this diverse landscape as we now have been presented for along this Master Thesis several times. The *explorability, stopability,* and *reachability* dimensions that are connected to traversability (Björneborn, 2017) are at the portal diminished and by design downsized. The portal does not encourage the users to explore the datasets in different stimulating fashions, as in the case with Lars, this can result in users searching, but not finding interesting data. Let us recall Lars experience of first at home searching the portal and then afterward at the workshop with this quote encountering the data visualization; *"A sub-cluster with '3d city model' that is interesting, why didn't I come across that when I looked myself? Have I just searched for the wrong terms?* Maybe Lars was just unluckily searched the wrong terms, but maybe the portal just did not encourage him to *reach for different resources* on the portal. At the workshop, he was encouraged to slow down and immerse while exploring the visualizations. Maybe Lars would have reached the same interesting datasets on the portal if he just was encouraged to slow down and immerse, maybe the explorability of the visualizations was a crucial aspect, but open

gov data it-in-self does, to my experiences, not explicitly facilitate *explorability, stopability,* and *reachability.* If returned to the example of the car driver, we can now add that even if the car is stopped and the landscape gazed, the landscape does not invite the car driver to explore and spend time here.

Sensoriability is described as *how rich affordances are there for sensing different resources in a given environment* (Björneborn, 2017). The visual exploration of open gov data in the workshop showed interesting unplanned data encounters when interacted with by the users. But other senses that the sight is confined by the digital existence of the technology. The data, when presented on the portal, was plain, simple, and very two dimensional with rows and columns. But an interesting aspect of open gov data, also in comparison to library material like books, is the tangible existence of the data, most of the open gov data is out there in the physical world, as a blueprint of a property of the city. When talking about the data with participants, it became much more real and tangible by relating the data to its blueprint form; the world out there. This experience is present, e.g., when Lars suddenly and enthusiastically finds data about Copenhagen trees, enthusiastic not because of the usability of the data but because of its tangible, sensible form. The sensoriability affordance is further referring to *contrasts* and guiding *pointers*. As earlier described, open gov data initiatives are highly focused on use cases as a strategy for communicating open gov data. This strategy can be persuaded as one big pointing exercise. The portal, on the other hand, is not designed with noticeable intentions to point the users in specific directions. Again, this might illustrate that the portal is intended to be a contrast-less and opinion-less data warehouse with the opinion generated externally, but are by doing so not only *downsizing* serendipity affordances but are actually *counteracting* them.

10 The agile vs. The waterfall

We are again situated in TANT-lab that still is designed for the workshop taking place. The lab has been arranged with printed material on the big table, with handwritten categories on the whiteboard and with interpretative visualizations visible on the big touch screen standing at the end of the table. I am seated at the table in front of my computer connected to the big screen and with Alex and Jacob sitting across the table just as they are, focused but smiling - without any devices in front of them.
Alex and Jacob are running a company called "sp8ces"[80] that puts companies looking for an office together with venues, that can be used as offices like hotel lobbies, cafes, and bars. Alex and Jacob participated in UrbanTech Challenge and actually won with their concept "sp8ces - Airbnb for office spaces"[81]. Their winning concept relies on open gov data from virk.dk that allow sp8ces to obtain company related data to secure an easier and more frictionless user experience. Sp8ces is building up a concept where the users each day can choose a new favorite location for their office.
After a bit of discussion in the lab, we did together decide to look deeper into the user-choosing-venue-process aspect of sp8ces.[82]

I am going to elaborate on the challenge of being a government instance that strives to be a data supplier for the private and the citizens. As we are going to explore, the role as data supplier is fostering new tensions in the self-perception as a public body. The public, as data supplier, is now not only serving society, it is now also, as my informants describe it "competing" with the private tech industries by being compared with the data giants like Google and Facebook. This competitive situation is not self-claimed by the government in their proclaiming of the open gov data project but is a side effect of providing products in a market where these giant private companies are ruling and are assembling the standards for the products in this market. And further changing the user's expectations and habits.

But what is this competition? The government is not striving to make any of their open gov data initiatives a competition between the public and private (their goal is just the opposite). But

[80] They changed the name to "Lobby."
[81] https://challenges.dk/da/node/526
[82] Author fieldnote, Sp8ces, 2019

when the users perceive the situation in the light of comparison, it is difficult not to accept the competitive situation, like Thomas Theorem advice (Bossen and Lauritsen, 2007), where concepts exist if people claim its existence. The discussion is now not whether such competition is desirable but *how* the public is competing with the private.

We were at TANT-lab discussion ideas that could support and enhance the users deciding process on the platform. We decided to systematically interact with each of the municipality visualizations to locate interesting topics that could be built into the sp8ces concept and support the user experience.
By first investigating the visualizations, then downloading a dataset found on the visualization and then uploading the dataset to the geospatial web mapping visualization tool Carto, we were able to visualize the base data from opendata.dk as geospatial data on a base map also containing the sp8ces venue's location. This data visualization journey left us with a two-dimensional map of Copenhagen, visualized with bike lanes and electric car charger data altogether with sp8ces venues location. Now, looking at this Carto map, I notice how Jacob and Alex' expression is shifting from the immediate excitement of visualizing the data into a mind of reflections. "I did not know that you could find data about the positions of the electric chargers in Copenhagen," Jacob tells and continues, "that is actually quite cool and may be useful if we have customers that are driving battery-driven cars."
Since this type of map visualization is more commonly being accessed at Google Maps, the comparison is obvious. "I don't know if you can find this kind of data at Google Maps, but the bike lanes, for example, can easily be found at Google Maps as well. The thing is that Google has its method to obtain and generate the same data as the municipality, but Google is not generating the data because of any political use. Google is probably producing this data because they have detected a need for the data. And talking about Google maps, this Carto map makes sense, but if I should find interesting data about the city I would immediately visit Google Maps instead and check out the city properties found in there". Alex is nothing and thinking. As Jacob is hesitating for a moment Alex continuous, "I think that Google Maps is more interesting, also because we are looking for city properties that can help us scale the concept to other cities and Google map contain all of that. When we use open gov data, we do not know if the data also can be found in other cities, but Google Maps got the same data for almost all cities." Here Jacob takes over again, "And another very intriguing aspect about Google Maps is that I have been working with Google

Maps API before, and I can quite easily get the data, where the data from opendata.dk looked a bit more inconvenience for me."
After speculating a bit, I felt that it was my time to contribute to the reflective space. "I totally get your points, but.." I heard myself, almost as an automated response, saying 'but,' just like I had to come up with a cleaver evaluation of why they were wrong as if I was trying to sell them this open gov data. I held back my salesman's speech and continued instead with a question. "...How do you consider this to be an inconvenience compared to google maps?".
We talked for a while about Google maps, Jacob explained to me how he admires the structure of Google Maps as a simple base map with a continuously almost infinite layers on top of it including an extremely wide arrange of real physical elements represented on the map, with restaurants, bars, roads, bike lanes, train stations, parks, shelters and so forth. A perfect 1:1 edition of the world to be accessed as data points through the Google Maps API. And Google Maps does not exclusively concern physical elements but also includes properties of the physical elements like opening hours and recommendations and even analysis about the elements like transportation time, road rush hours and working hours of the venues.[83]

Alex and Jacob' Google Maps fascination is here also a picture of their frustration about open gov data that from their perspective are struggling to follow the high standards of "free" private product, in this case, Google Maps.
There seems to be an inherent innovative user-driven approach in the way that Google is structured and managed (Steiber and Alänge, 2013) and built on a structure that animate iterations of the platform and data structure, where the users are encouraged to give feedback even when they are not themselves aware that they are providing feedback (Copeland and Savoia, 2011). This way of *doing* innovation with constant feedback loops is categorized as an *agile* method approach (Hannola et al., 2013). Public administration, on the other hand, is often admitted to being lead behind with what is known as the *waterfall* approach (Hanolla et al., 2013). Here the design process is linear going downwards like a waterfall, starting on the high table with the outlining of *requirements* and ends all they were down in the internal system in the *maintenance* phase of retaining the requirements from the top table. This waterfall model is often criticized for not being agile, iterative, and an innovative approach. A Weberian reading of bureaucracy (Kornberger et al., 2017) shows that public administration bureaucracy is aligned with the waterfall analogy where Weber argues that bureaucracy is administered with immense

[83] Author fieldnote, Sp8ces, 2019

persuasion on individual competencies. This foster a chain of bureaucracy commands, where the most individual competencies are giving order to the one under - like a waterfall[84]. Due to bureaucracy and other factors like; too general and broad societal concerns, data ethics, resources, and the lack of competitive attitude, it is very difficult for the public to perform the user's requirements. The same users are daily presented for user-friendly and well-structured data management when visiting the private tech giants not subject to public bureaucracy and general public concerns. The question is then, where and how the public sector is situating itself in this landscape among the private giants.

One may argue that this way of approaching open gov data in the scope of a competitive arms-race with private data giants simply is not the present case. The public sector would happily deliver open data to the private giants as well as any other private company or citizen. The public does not care if the open gov data users are also interested in other data than the open gov data. But this outlined competition is from my perspective not framed as normal market competition. No, the competitive situation between the public sector versus the private sector is a question of users expectations and anticipations for the products delivered to them. Big tech companies are pushing the standards for data access and data management both with the continuous development of computational power but even more important with user-friendly interfaces that keep increasing the user's expectations for products (Steiber and Alänge, 2013). The big shift is now that the increasing material standards not just are increased inexpensive product but also in products completely free of charge (money related!). This is due to a marked economic change in the web-based tech industry where the technical products no longer necessary are the source of income. Instead, the users are turned into products that produce pieces of information that the industry can profit out us, what Zuboff (2015) recognize as *the future market*. This marked change creates some sort of *Eldorado* of high-quality technology products that everyone "freely" can adapt and apply.

The question for open gov data development, as I suggest, is now how current and potential users of open gov data can be satisfied when one of the main advantages of open gov data, the *open* part, is becoming an expected standard in this Eldorado of high-quality technology products. This is the head of the nail in the public versus private competition as I see it; the user's high expectations for "free" products.

[84] Weber did not, to my knowledge, use the word "waterfall."

Discussion about the innovative role in the private sector and the public sector holds a long history and is a controversy that it-it-self is too comprehensive to outline here. But it is interesting enough to dig a few spades into the top sediment layers. Especially because of the obvious stooge to comprehend the private sector as colorful, innovative flowers and the public sector as greyscale bureaucratic rocks.

Hartley and colleagues (2013) discuss the innovative status and relationship between the private sector and the public sector. The main discussion they raise is skepticism against the prior statements that the public sector is less innovative than the private sector of which they state as a pure myth. They argue that both the private and the public are experiencing problems and barriers in the process of innovative work processes but these barriers are of such different character that they are incommensurable to compare. The situation between the private and the public is thereby very problematic to highlight with an innovative winner between the two.
Kornberger and colleagues (2017) are on the other hand, under the article headline; "When Bureaucracy Meets the Crowd," outlining the structural bureaucratically barriers in the public sector when dealing with open data. Kornberger and colleagues (2017) use a playground-analogy to show how it felt to work with public administered data where they translate the data into a park, that being open is a controlled playground. "*The park was, so to say, turned into a playground with the administration as guardian.*" (Kornberger et al., 2017) This playground-analogy is showing their frustrations about a bureaucracy-guarded playground, meaning that open gov data are supposed to be a creative playground but are controlled and stopped by the bureaucracy as the guard.

Google Maps is managed and structured with a strict focus on the *agile,* where user analysis and platform surveillance is a considerable part of the business (Copeland and Savoia, 2011). The public sector has, in this sense, restrictions that pose the bureaucracy and maybe blockades innovation, with *"secrecy, expert knowledge, written files, and rules*." (Kornberger et al., 2017) But at the same time take responsibility that the private sector is not obliged to in the same sense. (Hartley et al., 2013).

The important questions here is not only how the public perform as data supplier in the field of private tech giants, but maybe also how the public administration *reacts* and *reframes* it-self in this landscape. In the light of the Aktich and colleagues (2002a) inspired *interessement model* it

is simply not just the public open gov data products in-it-self that is a mean of innovation but instead the actors and the landscape around it that also have to be reformed and reshaped to successful innovation. That is why I am located at TANT-lab with Jacob and Alex from Sp8ces in an attempt to understand and aid reshaping both the open gov data portals *and* to understand the landscape it is located. Sp8ces have helped me understand the landscape by scoping open gov data projects and portals in the same scene as private tech giants. This opens new discussions about the open gov data projects existence justification and self-understanding. These open gov data platforms that by the government are scoped as a profitable and free resource that everyone simply can adopt is in the above-discussed light situated in a territory of high user expectations and predefined ideas of public as innovator versus the private sector.

10.1 Trust

Another striking aspect in the competitive situation between the government and the private tech giants is the concept of *trust*. Are open gov data trustworthy because it is produced by the government, or is it the other way around? It seems to be difficult to investigate if the users rely more on commercial data sources like Google and Facebook than the governmental, but it is not difficult to conclude that the participants of my workshops admire the products of the bigger tech companies in favor of the governmental as just shown above.

Trust relates both as the subject of trustworthy data and trustworthy data sources. The interesting aspects are *data trust* and *data quality*. Data trust is not just *trust* in the "correctness of the data" also trust in the maintenance of the data, the continuously of updates, and new issues. The data trust is also connected to the quality of the data. But quality depends mainly on the environment of what the data is supposed to be used for. Data quality is also highly related to the requirements that users of the end-product of open gov data are asking.

"But how can we know that the government is going to keep maintain the datasets, imagine that I built a whole company on top of some of this data, and then they decide to close the dataset down, who should I trust this data source?[85]
One of the participants asks me this highly relevant question. One striking aspect of this question is we trust Facebook as a data source more than the government? Facebook keeps

[85] Author fieldnote, diverse group, 2019

closing down their API's, and it seems like an unstable source for a business concept to be built upon.

Another participant at the workshop is highlighting in another context that the data he is working with does not need to be a hundred process "real." He argues that the data is interesting for him, and his research if he rather simply can access the data and play with it. Especially if he can incorporate the data into other data visualizations to interpret the data visualizations together with an issue expert or a citizen according to the project, this would create value for him. He is here arguing that the contents are not as important as the act of easily becoming familiar with the data. This questions the concept of data quality. Because what is data quality and to whom is the data of high quality or not?
Two participant at UrbanTech, e.g., was looking at the same dataset, one of them delighted that it was updated monthly, the other annoyed that it did not update daily. These aspects are important when building up a data trust culture.

The data trust culture is here referring to as the act of getting familiar with the data, the data sources, data platforms, and in general, the open gov data situation. One way of supporting the trust data culture is to structure the data in ways that allow the users to get familiar with the data, e.g., the data platform API. This will lift the trust to the data since it now exists in a familiar structure. The companies that I have been in contact with are all keen to constitute a trust relation to the data but before that can appear it is important to build a trust relationship between the actors; users, data, data sources, and data platform. But how can the data operationalize a trust relationship with the users? This is a further subject for discussion later in the discussion part.

11 Ethics - data, data and data

Just after presenting the lab, introducing myself, outlined the concept for the workshop and just after assuring Lars that the idea with the workshop beside my interest in his perception on open gov data is also to expand his knowledge about open gov data. "The workshop is attempted to be designed in order to give you insights about the machinery of open gov data platforms that hopefully leaves you with enhanced knowledge about open gov data in the end," I explain. "That sounds great" Lars reply. We talk a bit about his motivations for using data, earlier projects, and his previous role as an Urban planner that was frustrated about unstructured urban data. His current data project concept is about enhancing this structure. I try to guide the conversation in a data entry point direction and tried to make him talk about his first data encounter. "Okay. You participated in UrbanTech Challenge, you did very well and won the minor prize of 25.000 DKK. So what was your entry to open gov data in the challenge? Where did you start to look for data? And how did you proceed from there?" Lars thinks for a moment. Starts to answer but then guides himself in another direction as he stumbles across the public-private data relation. "Look, I have been working with this kind of data since my educational time as an urban planner. (not that long time ago). There is so much interesting data out there, especially when mixed with private data as we also are seeing a lot. But it is very important for me to express that I am quite worried about how the data are flourishing between the private and the public. I believe that we should have some clear regulations and some rules about what data usage and harvest that should be legal in the public space. The citizen should know what the data about them are being used for, even if they with personal identifications are not in the dataset. Data ethics in data usage is very important." Not expected Lars to talk about data use ethics in relation to his UrbanTech Challenge data concept about 3d models of the city. Caught off guard, I am here not succeeding to follow up on Lars' ethical consideration about the public-private data relation that Lars is highlighting, but are instead agreeing with him and then further asks him about his first data encounter.[86]

Lars' data concern is problematic concerning the open data ideal that all data should be open and the open data license that everyone should be given the power to use the data. The intriguing aspect of this concern, as I approach it, is the situated-situation where it appears out of an ethical context. It tells something about the atmosphere at the workshop that it is

[86] Author fieldnote, Lars Hartmann group, 2019

inclusive enough to facilitate these concerns, that never was a topic at UrbanTech Challenge. It also indicates different conceptualization of the term and ideology *open*. As we are going to elaborate a bit on, "openness" is here equilibrated with transparency, access, and the power of using the data.

Vestergaard (2019) is discussing the power relation between the *private*, *government,* and *citizens*. Vestergaard (2019) argues that those who have the data about the other, hold power over the other. He further contends that these three branches of power relations are uneven, where he locates that the government and the big private companies are administering and sharing data about the citizen. Vestergaard (2018) holds that the government is proceeding into, what he acknowledges as, *new citizen management* where the government is controlling the citizen with increasing data about the citizen. This is what he is a sliding slope argument predicts will lead to the dystopia, *the digital totalitarianism*.
Zuboff (2018) is outlining how the third party data brokers, are dealing with what she acknowledged as *the future market*, with the data brokers selling and profiting of advanced analyzed data about the citizen that the citizen is not even aware of and are never gaining access to.
Both Zuboffs (2018) and Vestergaards (2019) critical and ethical concerns are highly important to discuss concerning the increasing usage of data - also open gov data. And these discussions are also noticeable in this Master Thesis exactly because of the power relationship between the government, private and citizen that also are processed in this project. But as open gov data initiatives are keen to maintain, open gov data does not include personal sensitive data[87] and are therefore not a subject for ethical concerns. Lars, as we saw above, does not only include explicit personal data in this ethical concern, "*People should know what the data about them are being used for […]"* he also includes non-person sensitive data. *"*[…] *even if they with personal identifications are not in the dataset".*
The concern that Vestergaard (2019) discusses is mainly focused on people and data about people, so are Zuboff (2018). But in the isolated case of *open gov data*, the grand project for The Partnership is actually about displaying and exhibiting the government by inviting citizens and companies to benefit from this data about the government. Open gov data projects tend to focus on companies, but the data is somewhat open for the citizen as well. In this case, the power relation is maybe opposite of Vestergaards dystopia, and can maybe in a further extend

[87] http://www.opendata.dk/blog/data-taselv-bord-open-data-lab-aalborg-lanceret

be acknowledged as a source to overcome Vestergaards concern of the oppressed citizen? The third-party data brokers that Zuboff (2018) assigns as one of the key oppressing factors, is in the case of open gov data using data about the government instead of the citizen and are naturally using the data to benefit their own company. But examples have also shown that the third party data brokers are exhibiting the data about the government to support citizen knowledge about the government to be able to make democratic and lighted decisions, like "kend dit folketing"[88] and "DinGeo"[89].

The problem here is, of course, that the data about the government frequently also are data about the citizen. The government/citizen distinctions are thereby difficult exactly to outline. Clear distinctions are further being complicated in the landscape of data-driven power-relation discussions when scoping the citizens not as a coherent group but instead as divided groups with individual propositions, as Kornberger and colleagues (2017) are outlining in relation to the use of open gov data.

Ethical data concerns are an intriguing part of designing the future for open gov data and not aspects that easily can be overcome by proclaiming that open gov data does not include personal data. The democratization of data access is also an intriguing element of open gov data and maybe an object to balance the Vestergaard (2019) power relations between the government and the citizen. But the openness of data also includes the data to be used inexpedient in the public space, to follow Lars' concern. The crucial aspect is here to develop open gov data project in respect and consciousness of not building unethical barriers into the projects. But also to use open gov data to discuss related Smart City unethical data use and maybe actively use open gov data as a counterpoint to balance the relationship between citizen, government and private companies - in favor of the citizen.

[88] https://kendditfolketing.dk/
[89] https://www.dingeo.dk/

12 Who is the user?

All the open gov data that are being generated and distributed are addressing different targets and are motivated by different underlying motives. This rise the intriguing question; Who is the actual intended user of open gov data? But why is this important - can't we just issue the data and then be more ignorant who is using it?

The Scandinavian participatory design tradition and design thinking argue that user insights and knowledge are key affordances in product development and management (Robertson and Simonsen, 2013:2, Hyysalo et al., 2016). All open gov data projects I have read about and have been involved with, are firmly focusing on the involvement of users in their work with open gov data. These open gov data user involvements projects are striving to inform and educate the users about the potential of open gov data but are also seeking to obtain user insights and needs to issue data that the users are considering interesting. The latest (to my knowledge) and most comprehensive open gov data user experience project is organized by opendata.dk and facilitated and developed by the consultant company Seismonaut[90]. This project was extremely focused and interested in user knowledge. This user-centric approach is similar in other open gov data projects. *UrbanTech Challenge* was exclusively about users insights by forcing the users to think and rethink open gov data concepts[91]. *Data dialogue cafes*[92] invites users and data owner to the same room, *Single case entry points* are informing other users how to use open gov data, and so forth with user inclusive open gov data projects. But all the above-mentioned projects are by explicitly focuses on the users - naturally inviting users. The open gov data projects that I have not been invited to, and there is a lot, might not be similarly excited about user engagement. I have not come across a single data platform developers project or similar technicality oriented project that invited the users of open gov data to participate. When open gov data projects are outlining that they are interested in user insights, two themes of questions appear one theme about the *user* and one about *insights*. In the former, intriguing questions appears, who is the user? And, to *whom* is it the user? In the latter, what kind of insights are interesting? To whom is it interesting? And, how are the insights going to be used and implemented in the product?

[90] http://www.opendata.dk/sites/default/files/odaa/metodemanual_a4_final_-_131218.pdf
[91] https://challenges.dk/en/challenge/urbantech-challenge
[92] https://www.cityofodense.dk/da/open-data-arrangement/

One of the main actors in The Partnership is The Danish Business Authority. They are explicitly subjected to the project commitment of supporting the business life and is therefore implicitly framing the users of open gov data in a business context of companies. This framing of the user seems to be the general idea among open gov data projects that are involving users. I have not investigated how user insights are being used and implemented into the actual machinery of open gov data. But UrbanTech Challenge did not explicitly or actively use any user insights obtained from the Challenge other than producing a user case collection[93]. The interesting situation, as I perceive it, is then to question if user insights only are interesting and useful in relation to enhance the issuing of relevant data or if user insights also are interesting and useful in other open gov data enhancement aspects? I have in the previous chapter discussed the inherent logic of the platforms that are asking the data issuers and the users to interact in certain ways with the platform. I also have in a previous chapter discussed how the users of the open gov data platforms are interacting with the platform concerning serendipitous elements. Smith and Colleagues (2016) are further highlighting an innovative design focus technical circumstances for the development of open data projects to be adept. This shows that the platform itself and not solely the data on the platform are active actors in the process of locating interesting data. It is in this behalf both the platform and the structure of the data that needs to be curated according to the needs and insights of the users.

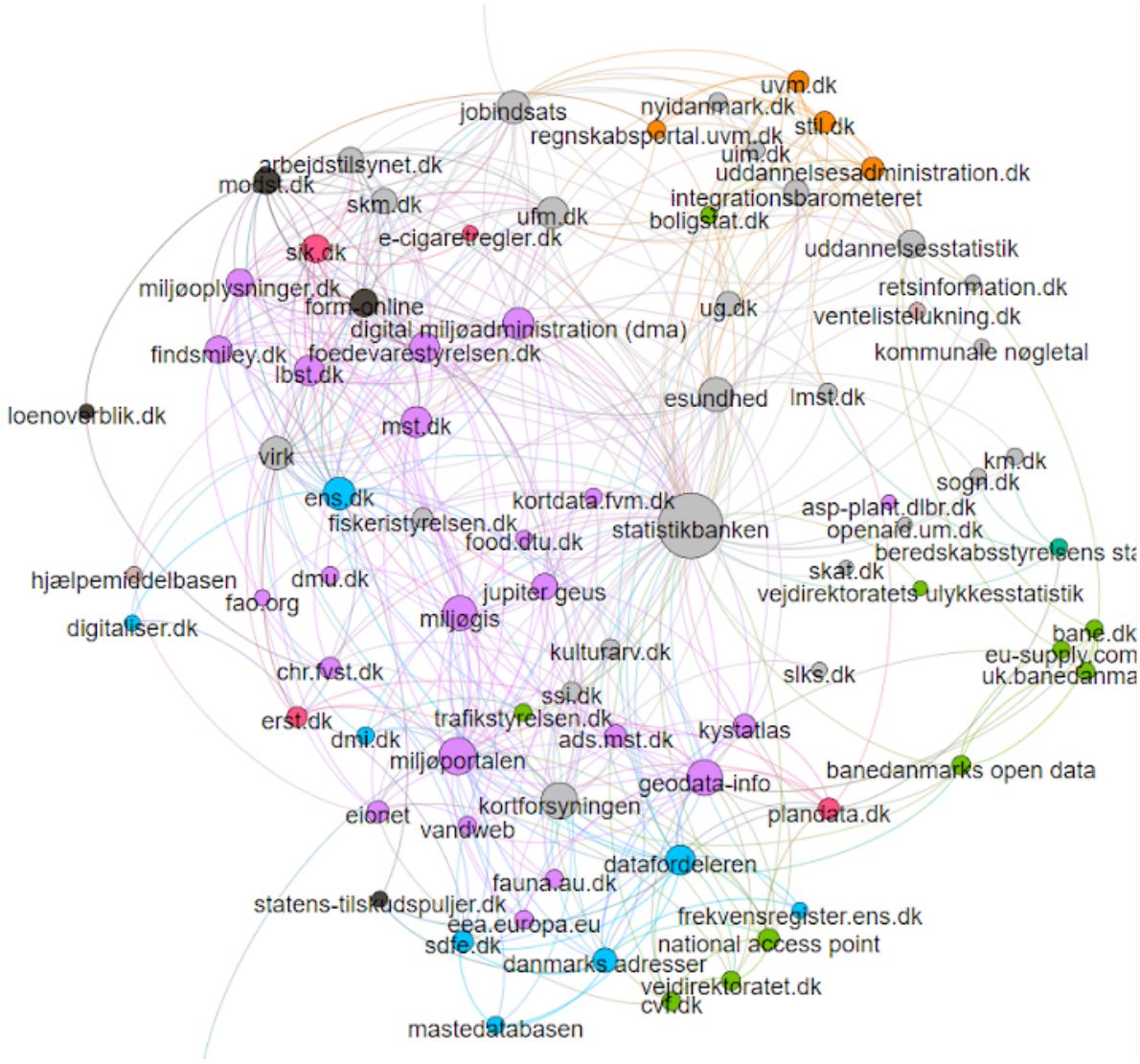
12.1 The 84 data platforms

I have earlier in details described the visualizations produced for this Master Thesis project. But besides the major visualizations, I also produced a visualization that is showing the connections between the 84 different data platform. In this visualization , the nodes are data portals, and they are connected with an edge if they share the same tags. This visualization shows how the portals are connected on basic of the contents of the portal[94].

At the TANT-lab, I am together with Jacob and Alex browsing through different open gov data platform. While discussion access and openness of different data platforms, I show them the visualization with all the 88 different open gov data platforms. "Who is the intended user of all

[93] Forthcoming
[94] The nodes are sized according to how many connecting that have and they are colored according to the ministry in charge of the portal.

these data platforms?" Jakob asks me. "The Partnership would argue that you are" I reply. Jacob continuous, "But then I do not understand why they make it so difficult for me to work with, with all these different platforms, accesses and structures. I really do not feel like the intended user of this product". Jacob looks a bit confused as he skims over all the different data platforms on the visualization. "But if they want me as a consumer of their data, why are they distributing the data across that many platforms and servers? If they wanted to be to use the data, maybe they should make it easier for me to access, and make me feel like the user." Jacob concludes.[95]



Visualization 12 - Open gov data platforms

On the workshop in TANT-lab, Jacob expressed that he did not feel like the target and intended user of open gov data. I had to ensure him that he and his company quite likely is the "third party

[95] Author fieldnotes, Sp8ces, 2019

developer" that the Partnership hopes to engage with the data; a young but data mature and data-driven startup that has not yet found the business foundation - still eager to find new data sources and new approaches for business. *"But if they want me as a consumer of their data, why are they distributing the data across that many platforms and servers?"* The answer to this question cannot be answered with user-friendliness but have to do with the variety and amount of data suppliers distributing the data. Jacob did not feel like a user of this product and was therefore not convinced or motivated to dive directly into this messy field of data on the different platforms, and this is a crucial point that we are returning to in the discussion chapter. Another interesting aspect of the discussion on who the intended user is appeared in an interview with Frans, who is working with opendata.dk, Copenhagen Solutions Lab and in the Administration for Technology and Environment in Copenhagen. Frans is involved in technical aspects concerning opendata.dk and is daily working with a variety of datasets on the portal at opendata.dk. Frans disagrees with the narrative that the companies are supposed to be the users of this data. Frans argues that the data that he daily manages on the opendata.dk portal, together with the employees in the administration for technology and environment, are foremost maintained, developed, and used internally within the administration. The user, in this case, is not the companies, not the citizen but the persons and the co-workers of the persons who issue the data.

Frans is welcoming me at his office at the Administration for Technology and Environment in Copenhagen. We talk about UrbanTech Challenge and the data visualizations that I have printed to show them to Frans. We are discussing the act of getting companies to use the data on opendata.dk. Basically, Frans did not sympathize with the idea that this data on opendata.dk is creating new innovative solutions abroad the municipality. "That is simply not how data works," Frans told me. "The data is extremely valuable for its purpose, and for what it has been generated on behalf of." Frans is very keen to explain that the data are generated with a purpose to solve a problem or a workload for the persons and departments that are administering the data.[96]

Frans is here expressing his perception of how the data uploaded to opendata.dk is supposed to be used. My perception of Frans dissatisfaction with third-party developers is related to the fact that he knows the motivations behind the development and generating of the datasets and thereby knows that the datasets are generated with a specific purpose. This knowledge is

[96] Author fieldnote, Frans - opendata.dk, 2018

telling the important story of data as fabricated data and not raw data since it is curated to solve a specific problem and not simply to contain objective amounts of data points. Frans is further deploying the argument that the data is intended to be used internally inside the building of where the data in the first place was developed. The argument here is rather simple, Frans does sympathize that the datasets are useful independent of the story behind it and the motivations for generating it. Frans believes that the open datasets distributed on opendata.dk are doing a great job of combining different departments inside the Administration but not necessarily outside it. Frans is here acknowledging another user of open gov data than The Partnership is flagging, namely an internal user within the walls of the Administration.

The focus on the users and user insights should not solely focus on the contents of data but in relation to Jacob' experience and Frans interpretations, incorporate a coherent discussion across the data platforms about who the intended users are, with the goal in mind to enhance the user's technical experiences on the platforms.

13 Analysis sum up

We have now been exploring the ecstatic busy urbanism of open gov data. We have been driving around investigating emerging topics that appear and have interacted with the digitized, physical, and social properties of the infrastructure to make it react. We are now looking for a calm place once again and are about to discuss what we have been experienced in the great and messy city of open gov data. We have experienced alternative approach for encountering the city, namely the socio-technical, driven by qualitative situations elicited by visual stimulations. This encounter fostered interesting discussions together with the users that in this project have been looking at the five analytic themes; *Data platform logic, Serendipity, High product standards, Data ethics, Trust"* and *The User*. These themes have been brought back from the city and are now the subject for further discussions and scrutinization. The motivations for encountering the technical and social material of the digitized city was encouraged by investigating the barriers for the usage of open gov data at non-governmental users. These barriers are now going to be explicitly examined and discussed.

14 Discussion

"Researchers crave flexibility, multiple pathways, and circuitous journeys. The image of the academic specialist, searching the shelves for a serendipitous connection, may seem quaint, but it remains powerful. The challenge for the digital library is to preserve this opportunity in cyberspace" (Foster and Ford, 2003)

We have now been introduced to a wide range of concepts and ideas. We have been guided through libraries, and we have been situated in TANT-lab with colorful visualizations on a huge screen. We have observed and discussed user concerns and frustrations. And we are now breathing out for the first time in a while. Because what is going on? We are now further trying to "slow down reasoning" and spend some time evaluating and discussing how the different concepts and topics above are related and relevant according to the research question. Enjoy this small breathing phase. The situation is soon going to become complicated as we are bringing the different topics, concepts, and concerns altogether. But first, I am going to break the first part of the research question into minor parts.

Research question; *"What are the primary socio-technical barriers for open government data to be adopted by non-governmental users?"*

We could start from behind with the non-governmental "users." The intriguing concept of participatory design is the focus on users, and since open gov data project strives to enhance the non-governmental use, this has been the focus for this project. The users and the usages are essential parts of the product and adaptations, which leaves us with the next concept "adopted." Akrich and colleagues (2002a) argues that adopting are related to adaptation. So an operationalization of scrutinizing adopting is to examine the user and technological adaptation by looking at *barriers*. The notion of barriers activates the major concern and drive in the research question, namely that something is blockading for this technology to be further adapted, used, or become a *common* if you prefer. This guides us in the direction of the actual usage and thereby back to the users.

The critical discussions that have been calling for attention during the project are in this regard; *Is the data supposed just to be out there - to be used, by someone, or is the data supposed to be curated and exhibited with a user in mind?* And, *what is the communicative role of the data platforms?* These are the main questions for scrutinization here. Further have other questions emerged, *is it a social material or a technical material that can deal with this concern?* And, *where does all this leave the government and their introspection?* And in the end, we are trying to discuss, *what the socio-technical barriers for user adoption of open government data are.*

14.1 Curation

Is the data supposed just to be out there to be used by someone or is the data supposed to be curated and exhibited with a user in mind?

The theme of *curation* continuously emerged in different forms during my work with open gov data. Curation is here deliberately curating something, like organizing or structuring with a goal in mind. The first encounter with *counter* curation was The Partnership informal and drowsy, slogan; *"we are generating the data anyway, why not try to have more people use it?".* The perception of open gov data as something valuable out there, somewhere that someone can use - as The Partnership propose is a quite unfocused and undirected strategy. This perception is a part of a figurative narrative where The Partnership is attempting to promote some sort of additional sales by fixating their attention at an unused potential that the concept open gov data holds. The slogan might be a figurative saying, but it is simultaneously becoming an inherent part of the promotion of open gov data. And it does not stop here. This perception on open gov data, as something fluffy but valuable out there, can be tracked down the *waterfall* from the manager table down to the structure, maintenance, and management of distribution in the data platforms. Let us look at this perception.

The Partnership is promoting open gov data as something generated *anyway* for another purpose as a residual product or a remaining stock. This residual promotion might financially be seductive, for the accountant, in the context of additional sales of a product already financed. And this residual product use might also talk into the narrative of a desired agile big data approach (Mayer-Schönberger and Cukier, 2013) of reusing and repurposing data just as the tech industry is doing.

Immediate concerns with the big data repurposing idea rise within the sphere of original usage of the data. Frans, as earlier presented, questions the reusability of the data. He targets the motivations for generating the data as a key aspect of the usability. Frans argues that the usability of data is interconnected to its reasons for existence. On this behalf, the data is not just a raw material that can be used in different situations. The data got a history that deliberately has to be considered in usage. Frans is here arguing that open gov data should not just be out there for anyone to use, as a residual product, but instead, be deliberately used for development. The exact *purpose of development* is, in this case, the *curated* use and user of open gov data. But beside staging the, to some, radical opinion that open gov data is a valuable resource, but only *internally,* inside the government, he is also suggesting that the data should be exhibited with a user in mind.

When researching how the structure, management, and distribution of open gov data are designed, the *residual product* or *remaining stock* perceptions on open gov data appears again. Broad arrange of entry points within multiple platforms, different data access possibilities, multiple API's, no data content overview, and in general no coherent design strategy. All this talks into the undirected idea of the open gov data as something out there, somewhere and can be used by anyone, an approach that could be acknowledged as the *residual product approach*.

Open gov data is still a relatively new field with a growing and emerging attention attracted to it. The administrative situation of framing the development of open gov data takes time and resources but a project like "Den fællesoffentlige digitaliseringsstrategi 2016-2020"[97] and "grunddata"[98] are by regimentation trying to make a technical development of the concept of open gov data. The data and data platform regimentation and uniformations are slowly but steadily being developed. But innovations are not just about technical development and transformation but, according to Akrich and colleagues (2002) to build robust alliances in the actor-network acknowledged as *the interessement model*. *"Innovation is the art of interesting an increasing number of allies who will make you stronger and stronger."* (Akrich et al., 2002a). The act of building alliances in this ANT perspective includes both human and non-human actors. We are, in this sense, encouraged to constitute an *interessement* that animates both human and non-human in open gov data projects to mobilize innovative development. The residual product approach can here be compared to the Akrich and colleagues (2002a) *diffusion model* in the

[97] https://digst.dk/strategier/digitaliseringsstrategien/
[98] http://grunddata.dk/

context of perceiving open gov data with the innovative drive and value in-itself. The curated approach can be compared with the Akrich and colleagues (2002a) *interessement model* where the innovative drive and value are added by actor-network and the active change of the surroundings. So let look at the building process of interesting allies by curating the human and non-human socio-technical development.

14.2 Being the user

The question under scrutinization is still, *Is the data just out there - to be used, (as we now acknowledge as the 'residual product approach' and relates to the diffusion model), or is the data supposed to be curated and exhibited with a user in mind? (as we acknowledge as the "curated approach" and related to the interessement model).*

As earlier showed some of the participants did not feel like the intended users of open gov data when encountered the data platforms, both in terms of the particular data platform structure and in relation to the multiple numbers of data platforms with different data entries. This is an intriguing user perspective aspect of the residual product/curated discussion. The user is here, after not feeling like being the intended user, dis-encourage to explore the possibilities of open gov data. This was present at the workshops, especially with Jacob and Alex. But how does this affect their actual work with open gov data? After all, they won UrbanTech Challenge by cleverly using open gov data. Good point. And yes, they did win UrbanTech Challenge, but they have not before or afterward been working hands-on with the data, even the data they propose in their winning concept is still kept unopened on the platform somewhere - not even downloaded. The question of curated user design is here also related to the actual use and encouragement to the practical use. Let us keep the user in mind and "double click" on the data platform opendata.dk.

The question of how active the data platforms should be, was earlier questioned as: *"Should open gov data platforms, be platforms for data to be uploaded on and succeeds by being a warehouse for data uploads? Or should the platforms play a more active communicative role and also be evaluated on their success of communicating the data to the users? It comes down to the distinctions between, storing, displaying, or curated exhibited the data on the data portals."*

This frames the residual product/curated discussion as a discussion of passive storage versus active communication in the data platforms. The answer to this concern is highly situated in the given situation; let me demonstrate.

The portal on opendata.dk has earlier has been described as a passive data storage-focused to facilitate simple convergent searches. This portal did not guide Lars in the desired direction. Lars was searching the portal to build his concept to UrbanTech Challenge. He did not find any dataset interesting for his concept - or any dataset that guided him in new concept directions either. But the same passive simple search portal was a powerful tool for data investigation when combined with external materials, in this case, the data visualizations. The data visualizations at the workshop contained the same dataset as opendata.dk but was curated, exhibited, and designed with serendipitous elements. When Lars and I was interpreting the visualization, new ideas and inspirations to concrete datasets emerged, that we then easily would search for at the portal and investigate in greater details. This illustrates the picture and suggests the argument that both passive and active data encounter are interesting to work with. This further indicates that divergent and convergent data encounter are useful. And finally, it suggests that the data platform no matter passive or active, convergent or divergent, deliberately should be designed in favor of the intended user. Let me further elaborate.

At the workshops, the data entry was re-facilitated in relation to accommodate the requirements of the users. During UrbanTech Challenge, the *concern of data contents* was manifested. The data visualizations were generated in order to accommodate this concern. The data visualizations were in a *graphic ideation* process experimented with, tested, refined and in the end, introduced at the workshop as a technical material to interact with and elicit the dialogue with. As showed at the workshop with Lars, these visualizations were actively used as a prototype and intervention for an alternative data entry here as a serendipitous data encounter. This data encounter opened up new data approaches with specific topics that afterward would be further investigated on the portal at opendata.dk.

This elaboration suggests, that deliberately design the data entry with the user in mind, enhances the users experience with data entry. In this example, the data was not just out there. The data was curated exhibited in continuation of the use. The *use* was here encapsulated in the workshop setting with a facilitator and the time to consolidate. The initial use and data

entry was situated to this workshop space and not necessarily to a general data entry. The example suggest a reconfiguration of the data entry with the user in focus and indicate that data platforms execute an initial role in the data entry that further can be stimulated when designed as *active* - as the example showed.

14.3 Trust and the platform logic

The next actor in the *interessement* allies network building is the concept *of trust*. Trust comes in different versions. I have earlier outlined trust in relation to open gov data as a trust to the continuation of the maintenance of the data and as a trust to data quality. Trust is an often mentioned concern when talking about open gov data. This is because of the uncertainty if the data are going to be financed in the future. The data untrustworthy argument is often, What if I build a business concept on top of this data and then they decided to stop issuing it? This is a relevant question and a question no one can answer or guarantee.
An important aspect of data trustworthiness is the ability to become familiar with the data and data entry. When looking at the relationship between maintenance, data quality, and familiarity as users concerns of utilizing the data, the question that emerges if whether a social material or a technical material can deal with this concern?
Let us look at the platforms and their designed inherent logic. The metadata questions that the platform asks the data issuer is on opendata.dk trying to accommodate these concerns with technical material. The data issuer is asked to answer how often the dataset is being updated, who the maintainer is, and how to contact the maintainer and datasets can be filtered according to last updates. The metadata can here be acknowledged as the technical material that is dealing with the concern of trust. The non-human technical material is here combined with the possibility for human contact and support, with the contact information on the maintainer of the specific dataset. This is interesting. This metadata with maintainers email, provides the link between the history of the data with the data maintainer, issuer, and usage, with the user. Frans argued that data and the data history are interconnected and had to be taking into account the usage. The metadata on the portal at opendata.dk can be acknowledged as a small step in the direction of combining this link between data and its history, especially also with the "detailed description" of the dataset. But again, the residual product/curated discussion appears, because the logic of the data platform is either just an extra help or specifically designed and curated with the user in mind. Let me show an example.

Situated with Jacob and Alex at TANT-lab, we discussed the opendata.dk API and how they could use it. We had just discussed the topic of data trust and especially the concern of continued maintenance and updates. I showed them the actual metadata related to this concern, and we, therefore, ended up exchanging ideas to technical solutions. We discussed how we with the API could make a small program that would filter the datasets on the portal for the frequency of updates. The discrepancy then appeared later, when we were looking at a dataset on the portal about toilets in Copenhagen[99]. When a human opened this dataset, they could immediately discover that the dataset is no longer maintained, it simply said in the description. But if the program, we just discussed, would access this dataset through the API and examine the maintenance status, it would conclude that the dataset is being updated, quarterly. This example illustrates that relatively good and clear communication between the dataset and its history can culminate in discrepancy when the data use and user are not specifically designed for. This suggests that even metadata is not just *out there*, as the residual product approach suggests, but should also be curated with a user in mind. It further illustrates another important issue with open gov data, namely that this data is being issued and maintained by a variety of people with different conceptions of the data usage in mind.

The multiplicity of issuer further results in an increasing variety of spellings and systems management. Maybe not a problem for a human to read and understand, but challenging for a computer. But the multiple issuers is also one of the strengths of open data, that allow a great volume of data to be issued and published. The substance of big data is often referred to as the three v's (later four v's[100] and even five v's[101]) of big data is volume, velocity, and variety (Mayer-Schönberger and Cukier, 2013). All three key concepts of big data are facilitated here at the portal at opendata.dk. Neither volume, velocity or variety is counteracting the curated conceptualization of the use or user. Big volume of varied data in real time as the big data empathize gives the management of the cites another dimension digital government service (Bertot and Choi, 2013) but it also requires new epistemological challenges in analyzing the data (Floridi, 2012) and as shown throughout the project also challenges the communication of the data as we are returning to now.

[99] https://portal.opendata.dk/dataset/toiletter visited 03.06.2019
[100] https://www.ibmbigdatahub.com/infographic/four-vs-big-data
[101] https://www.ibm.com/blogs/watson-health/the-5-vs-of-big-data/

14.4 Serendipity and communication

We have earlier, co-experienced Lars' journey on finding interesting data on the portal at opendata.dk. Lars did not find interesting data at first when *searched* the portal. Afterwards situated in the workshop space with *encountering* the same data on the data visualization Lars and I located interesting data that Lars could further dive into. This has earlier been categorized as a new *communication* of the same material applied with serendipity affordances. It is important to keep the exploratory and discoverability elements of the data visualization communication in mind. Which takes us back to the discussion of how active or passive the data communication should be. The exploratory data analysis approach, as I have tried to foster in this project with the visualizations, can maybe be categorized as a rather passive approach, despite the attempt to awake the data entry alive. If comparing the material communication with the library, it becomes evident that the library is activating the material by handpicking specific materials and highlight those. This approach is actually closer related to explanatory visual communication (Boy et al., 2015). This might be relevant in the active/passive discussion. Because this suggests that the data communication besides concentrating on serendipity affordances also could treat both passive and active data communication. The focus on the serendipity affordances, *diversifiability, traversability,* and *sensoriability* showed a concrete enhancement of the data accessibility and further suggested actual examples for enhancing the serendipity elements. The experimentation with serendipity affordances also fostered the argument that curated exhibited data mediation and communication activates not just the data but also the user. This, once again, suggest that the data platform should not deliberately brag about the *residual product approach* but should instead focus the attention on curated data exhibition.

Let us return to the connection between the library and open gov data and discuss data entry on a practical level. Libraries have in this project been presented as an interesting, and maybe a bit provocative, alternative for op gov data to get data entry inspiration from. We have seen how the library facilitates a diverse data entry with serendipitous designed spaces and data exhibitions.
The open gov data relation to libraries examines the data entry on a low-practical level, and questions the passiveness of the data platforms, that might succeed in gathering data but fails in communicating the data. This comparison suggests that data entry should focus on

exhibiting the data, not afraid of decision making and favoring data over other data. Opendata.dk is doing this[102], but not on the platform. Maybe this approach is enough to accommodate, externally, outside the platforms. The workshop was also an external accommodation of serendipity affordances. The crucial aspect is that the front page or the first data encounter should contain serendipitous affordances. If the portals are not designed with serendipitous affordances, then it maybe should not be the first data encounter the users are facing.

Another interesting aspect of library storage is the decision-making process of *where* to store to the material. Physical material can only be stored in one place. Libraries, therefore, had to invent systems that could accommodate this "physical problem" of storing. Digital data, on the other hand, are not subject to this determinism. Basically, "*there is no shelf"* as Shirky (2008) writes, and argues that the determinism of interpreting digital storage as physical storing is not just a fraud, it is preventing innovation. This argument opens new conceptualizations for data entries that further calls for alternative serendipitous data encounters.

While open gov data is trying to stabilize itself in the environment, libraries are continuously repurposing them-self. Because like anything else, libraries have to adapt to the surrounding natural development, where information, e.g., no longer are identical and equivalent to physical materials (Casey and Savastinuk, 2006). This changes the requirements for libraries that now have to rethink their bodily presence and their presence as a source for civic service (Bilandzic and Foth, 2013). Libraries are, besides distribution of materials, acknowledged for their role as mediators and facilitators for social activity (Gaus and Weech, 2008), community gathering meetups (Aabø and Audunson, 2012) and social learning and collaborations (Sinclair, 2007).
This challenges another perspective on open gov data initiatives when they also are compared with the multipurpose of the libraries. Given and Leckie (2003) further describes libraries as,
"The right to be in these spaces, to use them in certain ways, to invest them with a sense of ourselves and our communities—to claim them as ours and to be claimed in turn by them - make up a constantly changing public culture." (Given and Leckie, 2003)

This focus on the interconnectedness between the user and use suggest that the multipurpose and the user's ability to become a part of the library; both shapes the library and the user.

[102] https://www.linkedin.com/feed/update/urn:li:activity:6540151336304955392 visited, 05.06.2019

Translated into open gov data, this means that including the user by repurposing the usage of the platforms and initiatives together with the user, advances the adoption. This suggests that, besides activating the materials on the data portals with serendipitous inspired elements, the data platforms could also inspire new data usage by facilitating a multipurpose on the platforms – in order to keep or become, a source for civic service.

Where do these discussions leave the open gov data projects and development, and where does it leave the self-awareness of the government?

14.5 The government introspection

We have in the above discussions and analysis been presented for suggestions that favor the curated exhibition over the *residual product approach*. To sum up. The curated exhibition with innovative elements in the mobilization of *interessement* that actively modifies both the users and the product. And the *residual product approach* with the innovative driver in the *diffusion model* where data can just be placed out there where people then starts to use it. So, where does all this leave the government and their introspection?

The government is with open gov data distribution, as earlier outlined, challenged by the unwanted position as a competitor with big data drove companies like Google. The government itself is not the competitor, but the products they are offering and the data they are distributing are. I have earlier questioned if open gov data and its products sufficiently are interesting to adopt simply by the fact that it is governmentally administrated – or if it actually is the other way around. The comparison with Google Illustrates that open gov data are being situated in a territory of free product and high standards. This position can maybe be acknowledged as a success in itself, but it also implies that users not humbly are adopting open gov data isolated for their surroundings. This suggests that there is a pretentious battle going on in the field as data distributors. So, is the agile approach going to dominate the waterfall with its nimble footwork, or is the waterfall going to show its steady, powerful strength? No. it is now that kind of boxing fight that we are looking at here. The building of a strong network is an active act and can be violent but only figuratively. The key aspect of ANT inspired innovation is here switching to part two of the Ackrich and colleagues (2002a and 2002b) two-part articles with a focus on *strategic spokespersons* as the last topic of the discussion.

On a quite obvious and general level, the government has already appointed a spokesperson for open gov data, namely, The Partnership - whether it is strategic or not is another question. Focusing on The Partnership as the spokesperson allows the question on how to enhance the *strategic* part of the status as *strategic spokespersons.* And this spokesperson strategical-turn is here in conclusion of all the above discussions, acknowledged as a *curated management* strategical-turn. Maybe more as a recommendation than an analysis. I support that The Partnership in introspecting incur the responsibility of deliberately integrates the user (whoever that is) in the process of designing the prospects of open gov data. We have throughout the discussion chapter encountered the suggestion that a curated design with the user in focus enhances the usability of open gov data. Now, it is everyone's job to locate the use and that or those users in order for The Partnership to curate and design the future of open gov data as a future of active facilitation, deliberate usage of serendipity affordances, and in general abandon the *residual product approach* and actively favor the *curated exhibited approach*.

15 Conclusion

Welcome to the Smart city, to the ecstatic busy urbanism. Moreover, welcome to the *friction* of the Smart City; the open gov data platforms, the overload of entry points, the inadequacy of data overview, the overall undirected user design, the untrusted data, the ethical concerns, uncertainties, and the uninspired passive data storage.

Open gov data are no more than zeros and ones that merely are filling out spreadsheets and databases, without any opinions or attitudes to their existence - but only when perceived in isolation. However, when perceived as a socio-technical phenomenon in the landscape of political interest, data brokers and open ideologies, the picture is somehow more messy with the emergence of concerns and complications.
With a starting point at the ecstatic busy urbanism, this Master Thesis has been introducing, analyzing, and discussing open gov data aspect in order to locate barriers for the adoption of open gov data. The starting concept in this project was to scope open gov data as a socio-technical phenomenon in order to comprehend the social material and the technical material in the same sphere of exploration. By designing a workshop space with open gov data mediated through data visualizations interpretable by the participants, the theoretical conceptualization was methodologically operationalized. In this workshop space, and during the data visualizations interpretation, the participants were encouraged to articulate and discuss their experiences with and thoughts of open gov data. This data proximity workshop space was a qualitative situation with design object stimulations.
This space was the operationalization of the socio- (the qualitative situations with the participants and their experiences) and technical (the open gov data platforms and data mediated as data visualizations) phenomena. I have thought out the project introduced concepts and provoked discussion in order to approach open gov data with data and user proximity in focus.
The methodology of data-driven visual representations and interpretations allowed qualitative discussion concerning concrete hands-on data frustrations and barriers to the articulated and discussed. The data-driven qualitative setting was further mobilized as a mock-up. The visualizations were not only eliciting verbal communication in the design space; they also communicated the dataset with serendipitous affordances, where the participants were

confronted with new entries to encounter open gov data. This data encounter directed the participants into concrete data platforms and datasets, that further allowed frustrations, concerns, and efforts to be experienced and articulated - where the data barriers were expressed and manifested.

The barriers for open gov data have throughout this Master Thesis been mediated, showed, experienced, expressed and discussed as embodied concepts of the users, like frustrations, wonderings, despair, and wrinkled foreheads and eyebrows. However, the barriers were also mediated as exercised user actions like poor and uniform data usage, and unfamiliarities.

The most apparent barrier for open gov data to be adopted is with no competition the concern of data encounters, the uncertainty of where to find the data, acknowledge as *the concern of data contents and poor and uniform data usage*. The data have shown to be located across multiple platforms with different strategies for data entry, managed by different people with different concepts in mind. This has in this Master Thesis been recognized as *the logic of the platforms*. Data platforms are an interesting object for dealing with barriers. The barriers outlined are social material concerns mediated through the technical material on the data platform. The platform is facilitating the trust relationship between the user and the government and builds the data familiarity, namely, *the barrier of trust*. Data platforms are also subject to high user expectation with the users comparing big tech companies with the govern driven recognized as *the barrier of high expectations to the product*. The last data barrier is *concerning the uncertainty of a user* with *serendipity counteracting* aspects in the data communication.

When the above-outlined barriers on a structural level are approached, the most present and discussed open gov data barriers, in this project, is the structural approach of scoping open gov data as something, somewhere out there to someone, here acknowledged as *the residual product approach*.
This approach is present down the waterfall from the decision table at The Partnership of open gov data, to data initiatives, to the multiplicity of data entries, and down to the platforms and the communication of the datasets and even further down to the data maintenance. These rather significant and structural barriers are in contrast to the directed curation approach, in this Master Thesis acknowledged as *the data curation approach*.

These barriers have been experimented with, discussed and related to the data entry at the library which suggested that there is an urgent need for serendipitous affordances in the data encounter of open gov data, with actively curated and exhibited data.

These conclusions challenge how the data platform is communicating material and how the government in introspection is supposed to re-shaping it-self in order to accommodate this change of data approach from *the residual product approach* into the *data curation approach.*

Here in the company with the data curation approach is where I am going to leave this project, just with one last question to keep in mind.

How can open gov data initiatives be designed to overcome the outlined barriers and accommodate the user and the usage in the product? This is the intriguing question. This Master Thesis suggest that the answer should be something like; *open gov data should be designed to facilitate serendipitous affordances in the data entry and should be designed for usage rather than additional sales, - as 'the product' not the 'residual product'.*

18 References

Aabø, S., and Audunson, R., 2012. Use of library space and the library as place. Library & Information Science Research, 34(2), pp.138-149.

Abildgaard, M.S., 2018. My Whole Life in Telephones: Material Artifacts as Interview Elicitation Devices. International Journal of Qualitative Methods, 17(1), p.1609406918797795.

Abildgaard, M.S., Birkbak, A., Jensen, T.E., Madsen, A.K. and Munk, A.K., 2017. Playgrounding Techno-Anthropology. Easst Review, 36(2).

Akrich, M., Callon, M., and Latour, B., 2002a. Translated by Monaghan, A., The key to success in innovation part I: the art of interessement. International journal of innovation management, 6(02), pp.187-206.

Akrich, M., Callon, M., and Latour, B., 2002b. Translated by Monaghan, A., The key to success in innovation part II: The art of choosing good spokespersons. International journal of innovation management, 6(02), pp.207-225.

Bertot, J.C. and Choi, H., 2013, June. Big data and e-government: issues, policies, and recommendations. In Proceedings of the 14th Annual International Conference on Digital Government Research (pp. 1-10). ACM.

Bilandzic, M. and Foth, M., 2013. Libraries as coworking spaces: Understanding user motivations and perceived barriers to social learning. Library Hi Tech, 31(2), pp.254-273. https://search-proquest-com.zorac.aub.aau.dk/docview/1398001666?rfr_id=info%3Axri%2Fsid%3Aprimo

Birkbak, A. and Munk, A.K., 2017. Digitale metoder. Hans Reitzels Forlag.

Birkbak, A., Petersen, M.K. and Elgaard Jensen, T., 2015. Critical proximity as a methodological move in techno-anthropology. Techné: Research in Philosophy and Technology, 19(2), pp.266-290.

Björneborn, L., 2008. Serendipity dimensions and users' information behaviour in the physical library interface. Information Research, 13(4).

Björneborn, L., 2017. Three key affordances for serendipity: toward a framework connecting environmental and personal factors in serendipitous encounters. Journal of Documentation, 73(5), pp.1053-1081.

Blok, A., Carlsen, H.B., Jørgensen, T.B., Madsen, M.M., Ralund, S. and Pedersen, M.A., 2017. Stitching together the heterogeneous party: A complementary social data science experiment. Big Data & Society, 4(2), p.2053951717736337.

Boellstorff, T., 2008. An anthropologist explores the virtually human: Coming of age in Second Life.

Boy, J., Detienne, F. and Fekete, J.D., 2015, April. Storytelling in information visualizations: Does it engage users to explore data?. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (pp. 1449-1458). ACM.

Boyd, D. and Crawford, K., 2012. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. Information, communication & society, 15(5), pp.662-679.

Bossen, C. og Lauritsen, P., 2007. 6. symbolsk interaktionisme og STS. I: Jensen, C, Lauritsen, P. og Olesen, F. Introduktion til STS. Denmark: Hans Reitzels Forlag. s.139-156.

Brandt, E., 2006, August. Designing exploratory design games: a framework for participation in participatory design?. In Proceedings of the ninth conference on Participatory design: Expanding boundaries in design-Volume 1 (pp. 57-66). ACM.

Brandt, E., 2007. How tangible mock-ups support design collaboration. Knowledge, Technology & Policy, 20(3), pp.179-192.

Callon, M., 1984. Some elements of a sociology of translation: domestication of the scallops and the fishermen of St Brieuc Bay. The sociological review, 32(1_suppl), pp.196-233.

Casey, M.E. and Savastinuk, L.C., 2006. Service for the next-generation library. Library journal, 131(1), pp.40-42.

Chalmers, Rob. 2012 "Explainer: what is the open movement?" in The Conversation http://theconversation.com/explainer-what-is-the-open-movement-10308 accessed, 19.12.2018

Collier, J., 1957. Photography in anthropology: A report on two experiments. American anthropologist, 59(5), pp.858-859.

Copeland, P. and Savoia, A., 2011. Entrepreneurial innovation at Google.

Costa, S. (2019). My First 100 Days as a CKAN Contributor. [online] ckan. Available at: https://ckan.org/2019/04/03/my-first-100-days-as-a-ckan-contributor/ [Accessed 27 May 2019].

Crilly, N., Blackwell, A.F. and Clarkson, P.J., 2006. Graphic elicitation: using research diagrams as interview stimuli. Qualitative research, 6(3), pp.341-366.

Curran, J., 2013. Big data or 'big ethnographic data'? Positioning big data within the ethnographic space. In Ethnographic praxis in industry conference proceedings (Vol. 2013, No. 1, pp. 62-73).

Davies, T.G. and Bawa, Z.A., 2012. The promises and perils of open government data (OGD). The Journal of Community Informatics, 8(2), pp.1-6.

Docs.ckan.org. (2018). Tag Vocabularies — CKAN 2.8.2 documentation. [online] Available at: https://docs.ckan.org/en/2.8/maintaining/tag-vocabularies.html [Accessed 27 May 2019].

Europa.eu. (2018). European Commission - PRESS RELEASES - Press release - Digital Agenda: Turning government data into gold. [online] Available at: http://europa.eu/rapid/press-release_IP-11-1524_en.htm [Accessed 27 May 2019].

Floridi, L., 2012. Big data and their epistemological challenge. Philosophy & Technology, 25(4), pp.435-437.

Foster, A. and Ford, N., 2003. Serendipity and information seeking: an empirical study. Journal of documentation, 59(3), pp.321-340.

Gaur, A., Scotney, B., Parr, G. and McClean, S., 2015. Smart city architecture and its applications based on IoT. Procedia computer science, 52, pp.1089-1094.

Gaus, E.R. and Weech, T., 2008, May. The meeting room: Libraries as community centers for culturally diverse populations. In Providing Access to Information for Everyone: Proceedings of the 16th BOBCATSSS Symposium.

Geertz, C., 2008 (1973). Thick description: Toward an interpretive theory of culture. In The Cultural Geography Reader (pp. 41-51). Routledge.

Given, L.M. and Leckie, G.J., 2003. "Sweeping" the library: Mapping the social activity space of the public library. Library & Information Science Research, 25(4), pp.365-385.

Greyson, D., 2013, Information world mapping: A participatory, visual, elicitation activity for information practice interviews. In Proceedings of the 76th ASIS&T Annual Meeting: Beyond the Cloud: Rethinking Information Boundaries (p. 106). American Society for Information Science.

Guggenheim, M., Kraeftner, B. and Kroell, J., 2013. 'I don't know whether I need a further level of disaster': shifting media of sociology in the sandbox. Distinktion: Scandinavian Journal of Social Theory, 14(3), pp.284-304.

Guggenheim, M., Kräftner, B. and Kröll, J., 2018, Incubations: Inventing Preventive Assemblages. Inventing the Social, pp.65-93.

Hannola, L., Friman, J. and Niemimuukko, J., 2013. Application of agile methods in the innovation process. International Journal of Business Innovation and Research, 7(1), pp.84-98.

Hartley, J., Sørensen, E. and Torfing, J., 2013. Collaborative innovation: A viable alternative to market competition and organizational entrepreneurship. Public Administration Review, 73(6), pp.821-830.

Hey, A.J. and Trefethen, A.E., 2003. The data deluge: An e-science perspective.

Hine, C., 2000. Virtual ethnography. Sage.

Hogan, B., Carrasco, J.A. and Wellman, B., 2007. Visualizing personal networks: Working with participant-aided sociograms. Field Methods, 19(2), pp.116-144.

Hyysalo, S., Jensen, T.E. and Oudshoorn, N. eds., 2016. The new production of users: changing innovation collectives and involvement strategies. Chapter 1. Introduction to the new production of users. Routledge.

Jacomy, M., Venturini, T., Heymann, S. and Bastian, M., 2014. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. PloS one, 9(6), p.e98679.

Jensen, T.E., Madsen, A.K., Misfeldt, M., Munk, A.K. and Tamborg, A.L., 2017. Participatorisk Data Design: En ressource for Capacity Building. In Samskabelse og Capacity Building i den offentlige sektor (pp. 171-195). Hans Reitzels Forlag.

Kornberger, M., Meyer, R.E., Brandtner, C. and Höllerer, M.A., 2017. When bureaucracy meets the crowd: Studying "open government" in the Vienna City Administration. Organization Studies, 38(2), pp.179-200.

Kučera, J., Chlapek, D. and Mynarz, J., 2012. Czech ckan repository as case study in public sector data cataloging. Systémová Integrace, 19(2), pp.95-107.

Latour, B., 2005. Critical Distance or Critical Proximity. Unpublished manuscript. Available at http://www. bruno-latour. fr/sites/default/files/P-113-HARAWAY. pdf. Accessed March, 31, p.2014.

Law, J., 1992. Notes on the theory of the actor-network: Ordering, strategy, and heterogeneity. Systems practice, 5(4), pp.379-393.

Marres, N., 2017. Digital sociology: The reinvention of social research. John Wiley & Sons.

Madsen, A.K. and Munk, A.K., 2019. Experiments with a data-public: Moving digital methods into critical proximity with political practice. Big Data & Society, 6(1), p.2053951718825357.

Mattelmäki, T., 2006. Design probes. Chapter 3, Probes seeking human life. Aalto University.

Mathiassen, A.K. and Lindstrøm, M.B., 2015. Open Data DK skaber vækst og transparens. Geoforum Perspektiv, 14(25).

Mayer-Schönberger, V. and Cukier, K., 2013. Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt.

Morelli, N., De Götzen, A. and Simeone, L., 2019. Open data as new commons: design strategies to support the creation of communities and practices around a new shared resource. In Managing Complexity and Creating Innovation Through Design. Routledge.

Miller, D. and Slater, D., 2000. Internet. Harcourt, Incorporated.

Munk, A., Madsen, A. and Jacomy, M., Forthcoming. Thinking Through the Databody: On the Conditions of Inquiry in Digital Methods.

Munk, A.K., Meunier, A. and Venturini, T., 2016. Data sprints: A collaborative format in digital controversy mapping. Princeton, NJ: Princeton University Press.

Murthy, D., 2013. Ethnographic Research 2.0: The potentialities of emergent digital technologies for qualitative organizational research. Journal of Organizational Ethnography, 2(1), pp.23-36.

Nardi, B., 2010. My life as a night elf priest: An anthropological account of World of Warcraft. University of Michigan Press.

Open Group. (2018). Open Definition 2.1 - Open Definition - Defining Open in Open Data, Open Content and Open Knowledge. [online] Opendefinition.org. Available at: http://opendefinition.org/od/2.1/en/ [Accessed 1. Jun. 2019].

Opengovpartnership.org. (2019). Open Government Declaration | Open Government Partnership. [online] Available at: https://www.opengovpartnership.org/open-government-declaration [Accessed 27 May 2019].

Raetzsch, C., Pereira, G., Vestergaard, L.S. and Brynskov, M., 2019. Weaving seams with data: Conceptualizing City APIs as elements of infrastructures. Big Data & Society, 6(1), p.2053951719827619.

Robertson, T. and Simonsen, J., 2012. Participatory Design: an introduction. In Routledge international handbook of participatory design (pp. 21-38). Routledge.

Rogers, R., 2013. Digital methods. MIT press.

Rosenqvist, Tanja. and Heimdal, Elisabeth., 2011, January. The making of a mock-up: a story about how ideas are framed using reality as scaffold. In Participatory innovation conference (pp. 45-50).

Salganik, M.J., 2016. Bit by bit: social research in the digital age. Princeton University Press.

Shirky, C., 2007. Ontology is overrated—categories, links, and tags. http://shirky.com/writings/herecomeseverybody/ontology_overrated.html. Accessed 05.06.2019.

Sinclair, B., 2007. Commons 2.0: Library spaces designed for collaborative learning. Educause Quarterly, 30(4), p.4.

Smith, G., Ofe, H.A. and Sandberg, J., 2016, January. Digital service innovation from open data: exploring the value proposition of an open data marketplace. In 2016 49th Hawaii International Conference on System Sciences (HICSS) (pp. 1277-1286). IEEE.

Smith, M.L., Elder, L. and Emdon, H., 2011. Open development: a new theory for ICT4D. Information Technologies & International Development, 7(1), pp.pp-iii.

Sonnenwald, D.H., Wildemuth, B.S. and Harmon, G.L., 2001. A research method to investigate information seeking using the concept of Information Horizons: An example from a study of lower socio-economic students' information seeking behavior. The New Review of Information Behavior Research, 2, pp.65-86.

Stark, D., 2008. Searching questions: Inquiry, uncertainty, innovation. International Affairs, 1100, p.3355.

Steiber, A. and Alänge, S., 2013. A corporate system for continuous innovation: the case of Google Inc. European Journal of Innovation Management, 16(2), pp.243-264.

Sunstein, B.S. and Chiseri-Strater, E., 2011. Fieldworking: Reading and writing research. Chapter 5, Researching People: The collaborative listener. Macmillan.

Tantlab. (2018). Participatory Data Design. [online] Tantlab.aau.dk. Available at: https://www.tantlab.aau.dk/lab-philosophy/participatory-data-design/ [Accessed 7 Mar. 2019].

Taylor, E.B. and Horst, H.A., 2013, September. From street to satellite: Mixing methods to understand mobile money users. In Ethnographic praxis in industry conference proceedings (Vol. 2013, No. 1, pp. 88-102).

Venturini, T. and Latour, B., 2010. The social fabric: Digital traces and quali-quantitative methods. Proceedings of future en seine, 2009, pp.87-101.

Venturini, T., Munk, A. and Meunier, A., 2018. Data-Sprinting: a Public Approach to Digital Research.

Venturini, T., Jacomy, M. and Pereira, D., 2014. Visual network analysis. Sciences Pomedialab.

Vestergaard, M. 2019, digital totalitarianism,

Von Hippel, E., 1986. Lead users: a source of novel product concepts. Management science, 32(7), pp.791-805.

Winn, J., 2013. Open data and the academy: an evaluation of CKAN for research data management.

Zuboff, S., 2015. Big other: surveillance capitalism and the prospects of an information civilization. Journal of Information Technology, 30(1), pp.75-89.

Zuiderent-Jerak, T., 2015. Situated intervention: Sociological experiments in health care. MIT Press.

Zuiderent-Jerak, T. and Bruun Jensen, C., 2007. Editorial introduction: Unpacking 'intervention' in science and technology studies. Science as Culture, 16(3), pp.227-235.

Appendix 1 - Empirical sheet

name	arrangement	effort	Data proximity	Time	season
Opendata.dk frans	Interview	1	Data proximity	December	Winter
Datafordeler.dk	Interview	1	Data proximity	February	Winter
Luca Simione	Interview	1	No data	May	Spring
Lennart Björneborn	Interview	1	No data	May	Spring
Rentemestervej Library	Observation	2	No data	April	Spring
Sp8ces - Alex	Pre-Workshop	1.5	Data proximity	January	Winter
Sense analytics	Pre-Workshop	1.5	Data proximity	November	Autumn
CKAN platform	Presentation	1	Data proximity	April	Spring
CBS - startup	Presentation	1	No data	October	Autumn
seismonaut	UrbanTech Challenge	1	No data	November	Autumn
Danish Business Authority	UrbanTech Challenge	2	No data	Autumn	Autumn
Data owners	UrbanTech Challenge	2	No data	Autumn	Autumn
Partnership	UrbanTech Challenge	2	No data	Autumn	Autumn
UTC participants	UrbanTech Challenge	2	No data	Autumn	Autumn
Final event	UrbanTech Challenge	1	No data	December	Winter
Plandata	UrbanTech Challenge	1	No data	September	Autumn
Sp8ces - Jacob and Alex	Workshop	2	Data proximity	April	Spring
Lars hartmann group	Workshop	2	Data proximity	April	Spring
Academic group	Workshop	2	Data proximity	April	Spring

Appendix 2 - Fieldnote, the library

Even before entering the library and just standing in front of the only elegant building in the neighborhood a feeling of intended and designed enthusiasm emerged. This building holds the promise to offer me something, social or concrete, fun or serious, but the promise was very present while standing there in front of the impressive golden facade with a multicultural and diverse group of people bustling in and out of the, very public-alike, revolving door. (Author field notes, Library, 2019)

When entering the library numerous of things fought for my attention. A various of books looked at me all at the same time but from different places in this big open room. art constellations hanging and standing, colorful magazines, messy children, and librarians were all in a surprising calm vivar doing their own business. My purpose here was to find a quiet place with a soft seat for me, and a strong wifi for my computer. My intention was very clear I was here to study not to read books or magazines, not participate in any activity, or to rent a game, neither was my purpose to interact with anyone beside my computer and random procrastinations meet on the internet. I had a concrete seat in mind for my contemplation but decided to take a small detour through the library to get there. Overwhelmed by the calmness of the books I automatically slow down my movements. There was literally books everywhere. A bit disturbed by the fact that all book almost have the same dimensions I started to wondering who had "invented" this standard? Or from at techno-anthropological perspective, what socio-technical phenomena had influenced the closure of the standard of books. Probably something about human limitations in amount of concentrations, something technical and economical in the book printing phase, something cultural about where people read and how the books are carried, maybe even something to do with the size of bags as well. Whatever the reason being for the book size and the squared dimensions standard, must library shelf designers be thankful for the existence of it.
A rather simple, but very aesthetics, book storring design was leaving its marks on the library space as a whole. Rentemestervej library had decided to design the shelfs in a fashion that makes the shelves deep from one side and narrow form the opposite side, making it possible to store books on both sides of the shelf. This design choice determine how the books can be placed and thereby how the visual presentation of the books are being communicated. And here, surrounded by book, another reflection about books stroke me. It is not just the library that communicate the

books, the books themself, if given the change, are handling this commutative role as well. books are designed with about four outer interpretations: 1: The catchy front cover, that communicates the content of the book - modern books often includes a picture, illustration or thematic colours and the title and author name also in thematic fonts. 2: The insightful back cover, that communicates the book with a summary and reviews - often in the same visual theme as the front cover. 3: The informative spine, that in a brief and restrained fashion informs the name of the book and the name of the author. And 4: The measurable fore edge of the book, that in a solely quantitative manner gives the opportunity by simple eyeballing to compare and measure the size of the book (except that font size and margin is not calculated here). Another interesting aspect about books is the bodily sensorially presence of the book. After being found, or during the very process of finding a desirable book, the books at the library can be taken down from the shelf, touched and smelled, inspected and questioned all this without leaving the library.
Rentemestervej library had (by shelf design) decided to present and communicate their books both with the informative spine, pointing towards the user and also by the catchy front cover pointing towards the user, the former calls for a simple information search and latter calls for qualitative interpretations.
This library was indeed not a simple material warehouse or book storage space, this place was obviously designed for liveliness not for storring. But it did indeed as well store materials as just described. Books, magazine, and catalogues was consciously placed on shelves and stands in orderly organized systems. Note the plural "s" on systems. Because in here, multiple bibliometric system are build and orchestrated together. An alphabetic system, a numeric system, a thematic systems, an index systems, all are being orchestrated by the design of the library as the main conductor of this material-orchester. But the materials here are not "just" organized in bibliometric machine logic systems. The first shelf met when entering the library was dedicated to books that other users earlier had read. On the shelf in the middle of the big room are books that the librarian recommends. And even when sweeping around the bibliometric organized shelves stuffed with machine-logic, random selected books are presented with the front cover instead of the more usual space optimized book spine storing fashion. (Author field notes, April 2019)

After detouring the library space I found my intended seat and opened my computer. Hidden behind my screen I again started to look around, corrious of the activity taken place on the library. Feeling guilty by observing in secrete, I quickly found myself staring at a man in dialogue with a speaking touch screen half the size of himself - and he was no child. Their loud conversations

drowned the silence of the small room and everyone was listening. He was asking questions about the printer and the woman located somewhere remote (San Blas I hoped but probably not) was helping him with his technical request solely by speech. We were about twelve people in this open room all seated with headphones on, switching our attentions between books and computers. Some around big tables. Some in armchairs. The multipurpose huge touch screen facilitated questions for the remote librarian and general internet browsing but its main objective was to accommodate digital material searching, of which it showed the results an interesting fashion that I had never before seen. After searching, finding and resolving the material results, the big screen showed a map of the library with two dots; a you-are-here-dot, and a the-material-is-here-dot. Other ways of being guided around was though more traditional guiding methods, all kind of signs, headlines, arrows, pictures, and illustrations helped the users navigating the library. (Author field notes, Library, 2019)

Appendix 3 - opendata.dk visualizations

No 1 ***- parking, nature and tourism***

The biggest municipality cluster in encapsulated in with the themes, *parking, nature and tourism* which also are the three most used tags in the cluster. This cluster no. 1 is a rather big cluster that complicates the process of reducing the cluster and force it into a simplified conceptualization. But the spatialized cluster are often held together by a few bigger nodes that thereby are defining the cluster. This big cluster could have been separated into two clusters where "tourism" together with concepts like *kayak, fitness, sailing, toboggan run* and *outdoor* are forcing the cluster downwards and "parking" in opposite end with *shared cars, parking zones, roads* and *traffic* forcing it upwards in the two-dimensional space. The cluster is held together as one big cluster because of the node "nature" that are pulling the cluster together because of outdoor/mobilisation dataset like "ud-i-naturen" (translated, out-in-nature) that contains data both about nature experiences and the infrastructure around it with car parking, bike lanes and public toilets information. This cluster is therefore defined by the combination of the three aspects of nature experiences as the uniting force. Some of the other big nodes are, bike lanes, outdoor life, coordinates and fishing. A big cluster about *outdoor mobility and activity*.

No. 2 ***- technology, environment, maps and climate***

The second most volumeable cluster is concerning *technology, environment, maps and climate*. This cluster is held together by a common interest on environmental concerns. I did not manually color the cluster, so it is quite incidental that the color for this environmental cluster turned out to be green. Some of the bigger nodes are floating, coastline, ocean and cloudburst but also smaller sub-cluster appears with nodes connected to the big node, technology and environment, like a small sub-cluster of elderly concerns of health and services. A sub-cluster about environmental protections that unite the city planners to the cluster. And also a sub-cluster about grass and green areariers.

No 3 ***- traffic, mobilization and bikes.***

As with the first cluster are we here back to themes of transport and mobilizations just more focused on geodata and traffic counting. The bigger influential nodes are, *traffic, mobilization and bikes.* This cluster is quite united around traffic counting, especially bikes but also other

vehicles. There is a sub-cluster, in the bottom, exclusively concerning geodata. And a more integrated sub-cluster connected by *bikes* concerning lanes and routes.

No. 4 **- *citizen, age* and *sex***
The fourth biggest municipality cluster is concerning *citizen, age* and *sex* with big nodes like, citizen numbers, statistic, economy, election and population. This cluster is the least physical space oriented cluster, and are instead about society related concerns with socio-economics and demography in focus.

No. 5 **- *spare time, library, culture and sports***
This cluster is centered around the terms *spare time, library, culture and sports.* The cluster is faulty being held together with the sub-cluster in the upper right corner with the term *energy* and *heat* that unite the sports terms with electricity consumption terms. This is one of the issues of this type of text-based topologies data analysis. If the initial purpose for this project was to map clusters of open gov data should this types of misinterpretations had been fixed.

No. 6 - ***buildings, 3D models, city models* and *areas***.
The sixth biggest cluster is about, *buildings, 3D models, city models* and *areas*. Some of the bigger themes are conditions concerning building and building permissions but also other regulations like environmental related and heritages.

No. 7 **- *children, youth, education* and *schools.***
The cluster is about young social care, with topics like *children, youth, education* and *schools.* The cluster is mainly about schools related data like numbers of children and teachers at a school and also very much about recreation like leisure clubs and playgrounds.

No. 8 **- *real estate, housing and buildings***
This cluster about *real estate, housing and buildings*, is very much hold together by The building and housing register that are providing data about housing in Denmark.

No. 9 **- *bus, routes and public transportation*** *(Odder)*
This cluster is held together by a very coherent topic of bus-related informations mainly concerning public transportations issued at Odder municipality.

No. 10 ***– roads and road markings***

This narrow cluster about the *roads and road marking* is containing data related to the properties of the roads like containers, stripes and roadworks and maintenance.

No. 11 ***– environment, garbage, and gis***,

no. 12 ***– accommodations, activity, art, camping***,

no. 13 ***– case management, street lights, dogs,***

no. 14 - ***companies, employment, job,***

no. 15 ***– bikes, car traffic, smart city, eco,***

no. 16 - ***politics, elections, administration***

Appendix 4 - Rigsrevisionen visualizations

No. 1 ***- companies, management, social and tax***

This is the biggest ministry generated cluster. The cluster is held together by a very broad variety of company/business/finance/trade related themes in the top of the visualization. The bottom of the visualization is mainly about administration and management and the top and bottom are held together by common concepts of workload, employment and the labor market and the economics and administration in social counselling. Despite the size of this cluster is the topics represented rather homogeneous.

No. 2 ***- international, invest, airport and transportation***

Most of the clusters are highly determined by statistikbanken, but this second biggest cluster is more or less only data from statistikbanken. This cluster is statistics about international finance and transportation and not just as airport transportation, also transportation on trains, ships, roads. The cluster cover the transportations structural infrastructure. There is one major and quite dense cluster in the middle and smaller sub-cluster around it. One small sub-cluster in the bottom are about work accidents and speed limitations.

No. 3 ***- fish, forest, nature, mapping and protections***

This cluster is actually one of the few clusters that are being mobilised without any major influence from statistikbanken. The cluster about nature-related topics, are dealing with nature in a scope of inspections, protections, permissions, surveys and control. There is minor sub-cluster eg. one about water-related nature phenomenons like swimming and freshwater fish farming. And another sub-cluster about birds protections and bird areas.

No. 4 ***- education, school, student and criminal***

This cluster that maybe should have been separated into two cluster concerns two major topics one related to educational situations, like education, schools and students and the other major concern is related to statistics of criminal behaviour and offence. The two cluster are closely situated in the graph, but a closer examination reveals that the criminal cluster frequently uses terms like education and social conditions in their statistical comparison.

No. 5 ***- cash benefits, employment activation and sickness benefits***

This cluster is truly about social help and the social welfare system. The major topics are *cash benefits, employment activation and sickness benefits and so forth.* The data in this cluster is almost issued by "The Danish Agency for Labour Market and Recruitment" which makes perfect sense when examine the cluster, a quite homogeneous cluster indeed.

No. 6 ***- industry, farming, EU, production and pigs***

This cluster is about the farming industry with focus on the industry part of farming. This cluster is not about pets but on animals and farming in the scope of capital, regulations, antibiotics, federations and productions.

No. 7 ***- provinces, children, capital, household and residence***

Cluster number seven is about social housing conditions again from a statistical perspective. The cohesion in this cluster is a statistical examination and comparison of socio-economical properties like, income, housing, children, cars and assets.

No. 8 ***- ecology, economy, agriculture, fishing and business***

The statistical cluster cohesion is again present here, now just statistical comparison and perspectives about agriculture, fishing and ecology-economics.

No. 9 ***- professions, citizen, commuting, residence, churches and women***

No. 10 ***- national economy, economy, garbage and libraries.***

Acknowledgements

During my Academic life and this Master Thesis I have been in contact with so many great people that I would like to reach out to. A special thanks to my supervisor Anders Koed Madsen who is a great source of inspiration and exciting discussions. Thanks to TANT-lab for inviting me inside the machinery and introduced me to a world of critical data management *and academic amusement*. Thanks to Greentech Challenge for offering me responsibility during my academic internship and thanks to Aalborg University for teaching me nuanced criticism in a hyper complex world.

For this Master Thesis I would like to thank all the workshop participants, and all involved in the process of investigating open gov data. Thanks to:
Lars Hartmann. Alex and Jacob from Sp8ces. Asger from TANT-lab. Birgitte and Frans from opendata.dk. Jesper from The Danish Business Authority. Lennart Björneborn. Luca Simeone. Anders Koed Madsen. Rentemestervej library. Frederik, Anna, Caroline and Joachim from Greentech Challenge. Stefan from the Agency of digitisation. Christina from Plandata. Icelandic mountains. CBS startup. TANT-lab. UrbanTech Challenge participants. Data owners; Martin, Ane, Stig, Tom and Finn. Morten from Datafordeler.dk. Google for great products. Mathieu and friends for developing Gephi and in general thanks to open data and open source for being free and awesome. And many more.

And of course a big thanks to Anna.