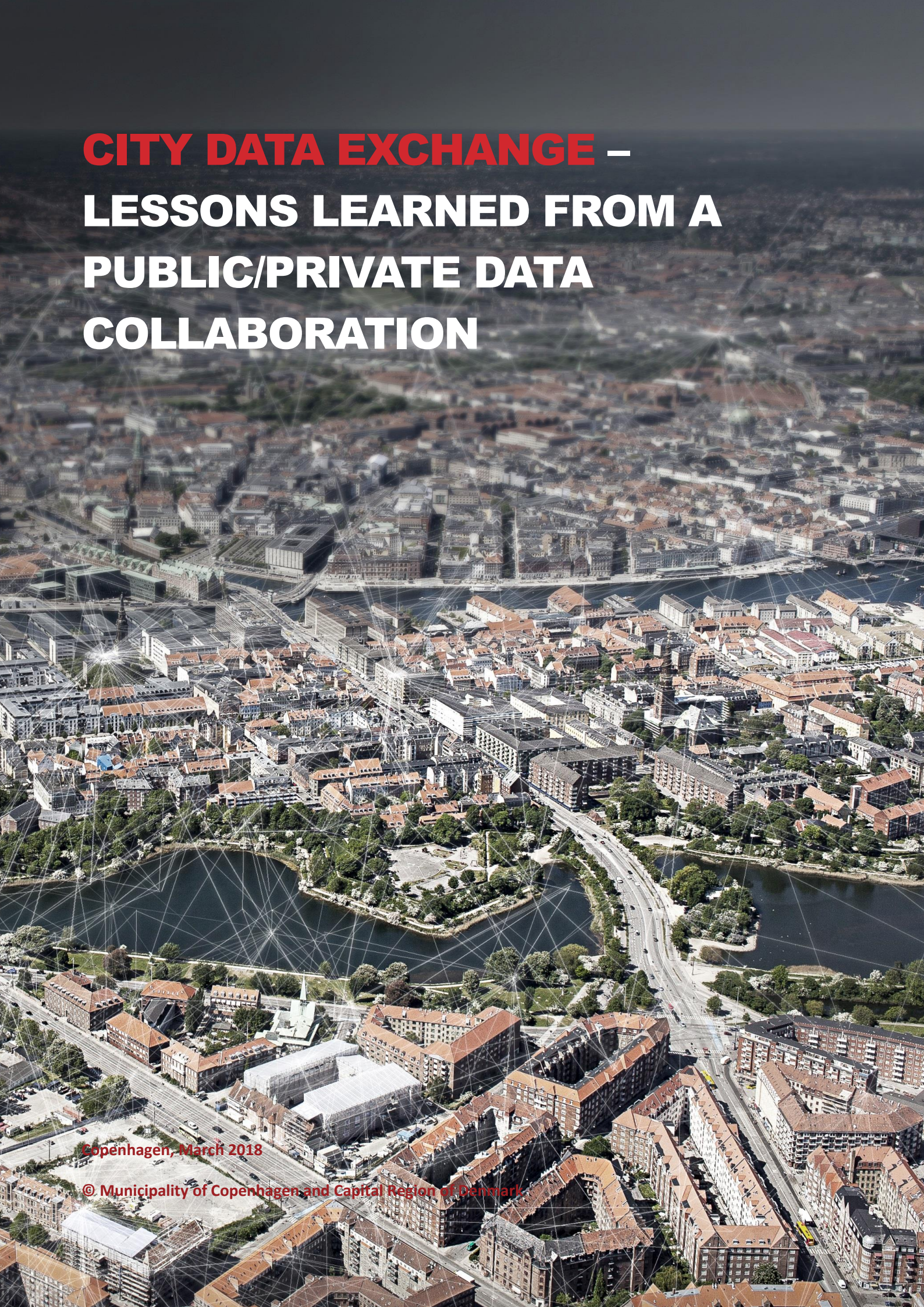
CITY DATA EXCHANGE – LESSONS LEARNED FROM A PUBLIC/PRIVATE DATA COLLABORATION



Copenhagen, March 2018

CONTENTS

SUMMARY

The City Data Exchange (CDE) is the product of a collaborative project between the Municipality of Copenhagen, the Capital Region of Denmark, and Hitachi. The purpose of the project is to examine the possibilities of creating a marketplace for the exchange of data between public and private organizations.

The CDE consists of three parts:

- A collaboration between the different partners on supply, and demand of specific data
- A platform for selling and purchasing data aimed at both public, and private organizations
- An effort to establish further experience in the field of data exchange between public, and private organizations

In 2013, the City of Copenhagen, and the Copenhagen Region decided to invest in the creation of a marketplace for the exchange of public, and private sector data. The initial investment was meant as a seed towards a self-sustained marketplace. This was an innovative approach to test the readiness of the market to deliver new data-sharing solutions.

The CDE is the result of a tender by the Municipality of Copenhagen and the Capital Region of Denmark in 2015.

Hitachi Consulting won the tender and has invested, and worked with the Municipality of Copenhagen, and the Capital Region of Denmark to establish an organization and a technical platform.

The City Data Exchange (CDE) has closed a gap in regional data infrastructure. Both public-and private sector organizations have used the CDE to gain insights into data use cases, new external data sources, GDPR issues, and to explore the value of their data. Before the CDE was launched, there were only a few options available to purchase or sell data.

The City and the Region of Copenhagen are utilizing the insights from the CDE project to improve their internal activities and to shape new policies. The lessons from the CDE also provide insights into a wider national infrastructure for effective data sharing. Based on the insights from approximately 1000 people that the CDE has been in contact with, the recommendations are:

- Start with the use case, as it is key to engage the data community that will use the data
- Create a data competence hub, where the data community can meet and get support
- Create simple standards and guidelines for data publishing

The following paper presents some of the key findings from our work with the CDE. It has been compiled by Smart City Insights on behalf of the partners of the City Data Exchange project.

THE DEMAND FOR DATA

In this chapter, we examine some trends and patterns in the demand for data. Generally, any given organization is searching for data for one of two reasons:

- As part of an innovation or new product development
- To increase efficiency in an existing process or product

The most sought-after data across different sectors seems to be very similar upon first glance. However, when looking at the actual requirements in more detail, one realizes that the demand for data is unique from organization to organization.

PEOPLE MOVEMENT PATTERNS

The most sought-after dataset includes information on how people move around in different places, and times in an area. This data is labeled *people movement patterns*.

This type of data is collected through several different sources, like cell phone tracking, wireless connection counting, camera image counting, traffic sensors, visual surveying, ticket purchases, and numerous others. Each of these has specific attributes, which sets specific limits to what the data can be used for. The table below provides a few examples of what different sectors want to use *people movement patterns*, and why they want to use them.

Transport	Municipality	Retail
Transport market shares	Traffic planning	Precise marketing
Route planning	Use of public spaces	Locating stores
Advertising	Flood planning	Location specific offerings
Increased efficiency	Health/safety	Identifying primary customer segments
	Tourism, culture, events	

Each sector, and each specific use case sets a specific requirement regarding data properties, such as granularity, sampling frequency, privacy etc.

SIMILAR – BUT DIFFERENT

From a non-data expert perspective, data on *people movement patterns* sounds relatively straightforward. The CDE team held conversations with Danish companies, which included supermarkets, Copenhagen Airport, public transport providers, urban architects, and different departments at the municipality of Copenhagen. During these discussions, parties were asked about how they expected to use the data. Resultantly, it became clear that they had very different needs and expectations of the actual information derived from the data. This was also the message that came from potential data providers, such as telco's, credit card companies, apps and start-ups working with drones, and satellite data users. As shown below, data demand differed in the following dimensions:

- Time (Real-time vs. Historic, Frequency of Updates)
- Geography (Public space, Building, Street, 100*100m Squares, Post-code, Region)
- Data delivery (Raw data, Excel files, API access, GIS maps, Dashboards)

Transport companies requested information on the number of people travelling between different geographical locations to understand their market share, but also to adjust their offerings.

Tourism organizations asked to get a better overview of the flow of tourists from different countries to inform industry on how to provide better services for tourists

Often the data requested was very specific, and was further not readily available. Resultantly, the exact quality, and type of data needed was often very expensive to deliver to a single customer. As such, the business case for purchasing this data was often difficult to justify.

One of the ways that the CDE tried to apply itself was to find ways to bundle the demand for data. However, as individual needs proved to be very different, it was difficult to find a one-size-fits-all solution, or to provide data at a suitable cost.

SOLVE MY CHALLENGES

There is an aspect that adds to the complexity of the data requirements by potential buyers. This is the fact that buyers require much more detailed data than solely the number of people moving around. This is because data is used in a context that is related to a specific use case.

This means that for the *people movement pattern data* to be relevant to a specific use case, it has to be joined either with more detailed information, or with other datasets. Below is a non-inclusive list of examples:

1) Data on people, such as: gender, age, education level, income level, nationality
2) Transport modes, such as: bike, bus, walking, metro, car, car sharing, train, including origin and destination
3) Origin, including: residents, locals, visitors, tourists
4) Other factors, such as: weather, events, train delays, accidents, and much more

The combination of 'raw' *people movement patterns,* and the additional information above needs to be combined in order to deliver on buyer demands.

The only way that this can be achieved is by delving deeper into different use cases. It is also essential to work with data providers, and those who request data to identify new, and innovative ways to deliver affordable, and useful data for both small and large consumers.

Generally, the demand for data is very specific and closely coupled with a specific product, or use-case.

The datasets offered by suppliers through the CDE were very specific as well. This is because these sets were collected for a specific reason, and were thus a by-product of a use-case, or product in the supplier organization.

For instance, when reviewing the datasets published by the Municipality of Copenhagen, it was very clear that they were collected as part of daily undertakings, and city maintenance. In several examples, this data was not optimal for projects, such as information apps for playgrounds, public toilets, and the like. These use cases require data that is updated at another frequency, or that contain other specific information.

BARRIERS TO EFFECTIVE DATA EXCHANGE

This chapter will focus on some of the barriers that the CDE tried to tackle. It does not necessarily look at issues that are discussed on a regular basis in articles and in the media, such as privacy, data standards, etc.

IMMATURE MARKET

Companies from the banking, insurance, retail, transport, tourism, and engineering sectors, among others, are already buying data from traditional sources like National Statistics, and the weather institute (DMI). They are also buying consumer data from specialized companies, such as Nielsen. Other clients are buying data from specialized companies that combine open and private data to make data products, such as Septima, and Geomatic.

The objective of the CDE project was to establish a channel that could be used by organizations, which were not specialized in data markets, like telco's, app developers, taxi companies, transport providers, and online booksellers.

Typically, an organization's budget for buying and selling data is limited, as this activity is often not part of its core business or strategy. Although companies and cities alike are eager to explore opportunities, they are not yet ready to scale or integrate data exchange activities in their business, and operating models.

LACK OF USE CASES

Several of the CDE client prospects asked for examples of good use cases that showed how traditional companies benefit from the sale or use of data, but there are only limited examples available. This is one of the main reasons behind the reluctance to invest significant resources in making data available for purchase or purchasing data made available by other parties.

From both a buyer, and seller perspective, the lack of use cases is a barrier. Organizations interested in selling data want examples of similar companies that have been successful in selling data. Likewise, prospective buyers want examples of how other parties have benefitted from buying data from the same sources.

Creating full use cases are costly for both data sellers and data buyers. Unfortunately, the complexity of demand (as exemplified by the *people movement pattern data* example) is holding companies and cities back from getting started with using data in the first place.

FRAGMENTED DATA LANDSCAPE

Some data is available via various open data portals from third party organizations, and selected private sector companies. Databases like opendata.dk (the national open data platform for public municipal datasets) and 'datafordeleren' (a national data platform for national basic data about Denmark and its citizens) have worked to compile data. Despite efforts to bundle data sources, both small and larger companies typically ended up using the limited data they could find through well-known sources. While many clients tried to explore and find new information, they quickly ran out of both resources, and patience. This situation has been exacerbated throughout the past three years, as many more data sources have arisen. Further, as private sector companies move to sell their data, the fragmentation issue only continues to increase.

RELUCTANCE TO SHARE DATA ON AN OPEN DATA PORTAL

Selling data, and especially personalized data is still new to many organizations. There are several reasons for not wanting to share data on an open portal despite it being aggregated and made anonymous

One reason is related to data ethics. For instance, *people movement patterns data* is often created through customers' use of companies' products and services (cell phones as an example) Although complying with legislation, companies are concerned about the reaction of the media, and their customers.

In a recent case in Denmark, the telecommunications company TDC sold data to Visit Aarhus. This quickly sparked negative attention from the media. Many other companies have had the same, or similar issues. These companies mentioned that positive use cases, which outline how data has helped an external buyer to create a product, or service that would benefit consumers, and citizens in general, would help immensely moving forward. These companies want to thus sell data on a 1-1 basis so that they are able to ensure a good story surrounding the use case.

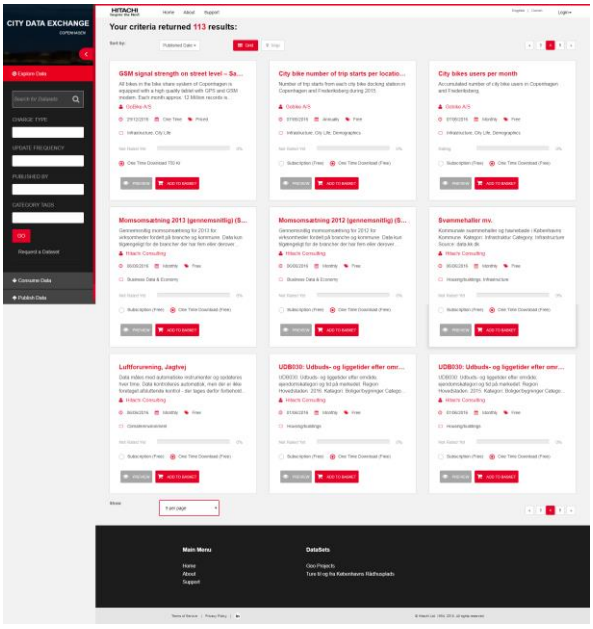
Another reason that clients are reluctant to put their data on an open platform is that their competitors can gain access to their data. Transport companies are reluctant to share their data on an open data platform for this very reason. This is the same for other types of companies, as well. Although it is easy to limit access of sales directly to competitors, it is difficult to keep other companies from buying data and reselling it to a competitor.

SKILLS AND COMPETENCES

Companies and cities are generally lacking the skills needed to work with different types of data from various sources. This is a barrier for many parties who want to sell data, as they do not have the skills to collect and prepare the data for the buyers, which would thereby allow them to transfer 'raw data' into information. The lack of insights into GDPR is also a barrier. Investing in data scientists, and equipment is expensive, and as there is a lack of use cases, many companies and organizations are holding back in making these investments. This makes it difficult to enter the market.

THE ROLE OF A TECHNICAL PLATFORM

One of the products of the CDE project is an IT solution for displaying, selling, and purchasing data. The platform was named the City Data Exchange. It was released in several versions throughout the project lifespan.



It includes the abilities to upload datasets for sale, to identify relevant datasets based on a series of criteria, to see metadata, to sample data from datasets, and to purchase datasets.

As a result of some of the barriers mentioned above, the exchange platform has had a fairly low impact on the project.

The exchange of data between seller and purchaser is nearly always based on an individual deal, which is suited for specific needs. As such, the need for a general platform has been less relevant in the CDE-project.

FUTURE TRENDS FOR DATA EXCHANGES

The insights from the CDE project provide a look into what can be expected in the future. These include the unknown future role of data platforms, the rise of data communities and collaboratives, data broker functionalities, more tools for easy data search, and data handling.

RISE OF DATA COMMUNITIES AND DATA COLLABORATIVES

Both potential sellers and buyers mentioned that a community to discuss opportunities would be highly beneficial. They requested a communications forum that would bring together data sellers and data buyers.

Most organizations stated that while data platforms are a good way to display sample data, many other services and tools would be needed for sales purposes. As previously mentioned, the reluctance of organizations to publish data on an open data platform is due to issues of data ethics, and competitors being able to access their data.

One of the growing trends is data collaboratives, where different data owners (public, private, IoT solutions) are gathered in one place. Here, interested buyers can request data and solutions. See this link for an example http://datacollaboratives.org

Based on the experiences from the CDE, we expect community platforms and data collaboratives to play a vital role in the future. These collaboratives will serve to co-create use cases with a wider community, and to identify which data is suited to solve a given problem.

DATA BROKER FUNCTIONALITIES

As the market for data as a product is still somewhat immature, several organizations have requested support on where to find information (data), and information on who could be interested in purchasing their data. Thus, there is a need in the market for different types of data broker functionalities. One possibility is through a data collaborative. This demand for support by the private sector will most likely increase as the market matures.

Many organizations do not have the needed tools or skills to handle data – either internally or externally. There is a demand for cost-effective, and easy-to-use tools to support the handling and visualization of information. This includes data displayed on maps, dashboards, and charts. Several organizations mentioned that, in the future, tools to handle various data types from multiple sources would be needed, as well as intelligent tools that can make quick and easy searches for data over many open data portals.

CREATING AN EFFECTIVE DATA INFRASTUCTURE

Work on the CDE provided insight into demand from the wider data ecosystem. It has shed light on the possibility of a wider data-sharing environment with data suppliers and data consumers.

Work on the CDE generated a lot of information on what is needed to make data sharing more effective. The CDE suggests the following three initiatives as a start to address the different issues discussed in these topic papers. These are based on findings from the past three years.

1) ESTABLISH SOLID USE CASES

Full-scale use cases that show how data from private and/or public sector can be used to address major challenges and/or provide new and innovative opportunities were the most requested item across all sectors. These use cases need to be based on either a specific business opportunity, or a challenge. The benefit of making a relevant full-scale use case is that it will not only indicate the project advantage, but it will also create insights into gaps in the process. It also will identify tools needed to support effective use of data exchange.

- Establish the business case– focus on getting both a consumer, and supplier perspective
- Be clear on risks and pitfalls, and how they can be avoided

2) CREATE A REGIONAL AND/OR NATIONAL DATA COMMUNITY

There are still a lot of unanswered questions, and there is a willingness to discuss opportunities for data exchange across companies and organizations. As such, there is a need for a common platform to facilitate discussions. This platform is also needed to begin to create use cases and tools to show the value of exchanging data. Both in Denmark and internationally, organizations are still trying to find the winning formula that will make data exchange between multiple stakeholders attractive and easy.

- Create a data-sharing community, which is a place to meet and explore opportunities
- Link to other activities and data sources, both national and international
- Identify data demand patterns across the public and private sectors

3) ESTABLISH COMMON STANDARDS FOR DATA SHARING

There is a good deal of work being done to find common standards for data sharing. However, these standards have not yet been implemented. Many of the cases in the CDE project surrounded the question of the format of different datasets. The requirements differ across sectors and organizations. Regardless, there is a need to investigate this and to promote these requirements. Based on the experiences from the CDE, we recommend starting with the use cases and one sector to investigate what standards are needed at a very basic level.

- Multiple sources demand tools to handle, and visualize data
- Limited need for visualization for each individual source – larger need when all sources are collected
- Create guidelines for data sharing, IT security, and privacy issues

There is a great opportunity for the Copenhagen Region and Denmark to build on the momentum, and insights from the CDE. Denmark has the right size to develop a model that will work. Many companies, and organizations have been part of the CDE journey, of which some are even selling data to each other today. There is a need to create a larger data-sharing community and to work on the use cases.

PROJECT ACTIVITIES AND RESULTS

Since the signing of the contract in April 2015, the CDE team has been in contact with almost 1000 people in the co-creation and promotion of the CDE. Hitachi, in partnership with the City and the Region, has carried out numerous activities with the aim of engaging with and understanding the data community. The activities included:

- 2 major events at the Town Hall
- 10 Info-meetings (introduction to the CDE, use cases, matchmaking, GDPR)
- 4 thematic workshops (energy, transport, tourism, city planning)
- Company survey on data supply and demand
- 30+ presentations in local and international events
- 300+ 1-1 meetings
- 10+ company workshops (explore data opportunities with individual companies)
- Monthly project meetings between Hitachi, the city, region and CLEAN
- 5 articles
- A data exchange platform
- 2 use apps.

The CDE brought together companies and organizations to discuss data-sharing opportunities. Many companies subsequently met on a one-on-one basis after these events. The CDE supported many meetings in a data broker function to understand the data ecosystem.

The CDE platform contains a few sets of priced data from various sources. Several purchases were made during the project period. However, these were not enough to make a sustainable business case.

MATCHMAKER AND DATA BROKER

The workshops, info-meetings and one-on-one meetings showed that there is a gap in the data exchange infrastructure. Companies, and public-sector actors are looking for a place to get support on how to work with external data, either from a buyer or seller perspective.

The CDE created a space for data-driven innovation projects and insights by bringing companies from different sectors, and different sizes together to discuss and explore the topic of a data exchange. The CDE team has acted as a matchmaker for many organizations who were given a chance to discuss data sharing with new potential customers and/or partners.

"The CDE gave us the opportunity to display our pollution data to a wider audience. We participated in several workshops, where we met potential customers. We had a close relationship with the CDE, and they have put us in contact with a number of potential customers inside and outside Denmark" Mathieu Carlier, CEO & Co-founder, Everimpact

From a data supplier perspective, the CDE team supported local companies in exploring the value of their data by identifying potential customers. For data consumers, the CDE team acted as a data broker by looking into challenges and opportunities and by matching their needs with data sources from both the public and private sectors.

One very important finding is that both the public and private sector are looking for the same types of information/data. This can be used for future data exchange developments, and provides many opportunities going forward.

ATTRACTING INTERNATIONAL ATTENTION TO THE REGION

In addition to promoting Copenhagen as an innovative region for smart city and data-driven innovation through international events/presentations, and media, the CDE has attracted international attention from both small and large organizations, and cities/nations globally. This has resulted in several high-level visits from companies. It has also sparked public-sector interest from countries, such as Singapore, Japan, UK, US, France, Spain, and Poland. These visitors also met with other CDE users during their stay in Denmark.

The CDE also caught the attention of international digital start-ups located in the US, Finland and France. These companies were looking to use the data from the CDE to work on new digital solutions, and to make use cases. These companies visited Copenhagen, and there is on-going dialogue. However, some important barriers still need to be addressed.

"In the past 2 years, we have met with the CDE team, who provided valuable insights into how companies are working with data. One important thing mentioned by the CDE team was the need to visualize the data. At BLOXHUB, we will have several facilities that support exactly that. It has helped us in making sure that we are focusing on the right things. Some of our members from outside Denmark are very interested in the CDE, as they see this as an opportunity to enter the Danish market." Torben Klitgaard, Director for BLOXHUB.

STUDENT ENGAGEMENT

The CDE also engaged with local universities. They gave lectures, and assignments at the IT University in Copenhagen. In return, this provided input to the CDE team, but also gave students insight into data exchange activities. Hitachi Consulting also invited a few students from the IT University to work with its technical team in London, as well as with the CDE team at Hitachi Consulting in Copenhagen.

NEW KNOWLEDGE FOR POLICY DEVELOPMENT

The engagement with local and international companies has provided knowledge about the wider data ecosystem and data infrastructure. The CDE team has worked closely with the city and the region of Copenhagen, which has provided insights into how to structure new data-based projects. The CDE team has also provided input for regional and national strategy creation around data sharing. It has also worked with the Danish Business Authority, the Danish Digitization Authority, Environmental Agency, DMI, the Danish Environmental Data Portal, and many more. This input has sparked new discussions on how to increase the uptake of data, including the creation of a better data-sharing infrastructure.

Jacob Laurvigen, Co-Founder of dexi.io explains the importance of a trusted data exchange like the CDE. "We have been partnering with Hitachi and the CDE from the very start, and it is clear that a secure and managed data exchange gives the comfort and trust that makes the transition to a data driven future more accessible." Jacob Laurvigen explains further, "There is no doubt that the CDE will fast track cities and companies, giving them a clear competitive and innovative advantage in the years to come. But it is critical that local governments and legislators back initiatives like the CDE. With the CDE, Copenhagen now has the chance to become a frontrunner in the Data Revolution."

SELLING DATA

Within its first year, the CDE team managed to gather more than 140 different datasets on the CDE platform. Many of these were datasets from various other open data platforms. They also included priced data from the private sector from organizations like the Danish Technological Institute, Saxo, GoBike, EverImpact, and Dexi.io. Data was also sold to several customers.

REFLECTIONS OF THE CDE IN THE CITY AND REGION?

CITY OF COPENHAGEN

The city is engaging with a range of companies and interest groups to explore further data sharing through Copenhagen Solutions Lab (https://cphsolutionslab.dk), and through membership of BloxHub, a co-working space and community designed to shape better cities. This is located in central Copenhagen, where the use of data is high on the agenda.

The focus on use cases and wider dialogue is also taking place in the city's involvement in Open Data DK (http://www.opendata.dk/) – a national open data platform for regions and municipalities. Here, public authorities publish datasets about the city, companies, the environment, and more. The platform is constantly being developed, and insights from the CDE are used to create value on how to ensure greater uptake, and usage of the data from the platform.

The city is looking into using external data, and is using data broker insights from the CDE to find new solutions to some of their most complex challenges. These include parking, pollution, and climate adaption projects.

As highlighted in the recommendations, visualization of data is important. Users of data, including the City of Copenhagen, are looking into how to combine data from several sources to develop tools to handle, combine and visualize data. Partnering with Cisco, the city has engaged in a new project, "Cisco Kinetic", where combining and visualizing is at the core.

The insights from work with the CDE have strengthened the focus, and skills of the city of Copenhagen, and its partners. This CDE has provided important input to the new digital era in Greater Copenhagen, where solutions to problems are now increasingly being based upon data.

THE CAPITAL REGION OF DENMARK

The Capital Region of Denmark has also joined the Open Data DK community. Like the city, it has used knowledge from the CDE in this partnership.

The Region has placed more focus on the importance of building relations and creating communities. Rather than solely looking at the technical aspects of a data platform, it has also considered needs and challenges. This is also evident in the regional growth and development strategy, and action plans. An initiative is set to create a common data hub, and community in Greater Copenhagen.

The CDE findings have shown that it is extremely important to build up the competencies of both data providers, and data users. This is now a large focal point for the region.

The CDE has worked with several companies from the telecom industry, and other sectors where personal data and data privacy are considered to be of high value. The Region took the findings from the CDE into consideration while developing Smart-City Cyber-Security Lab, which was established in 2017. It is also using cases from challenges working with telecom industry data to develop knowledge on how to move forward on the data agenda.

Finally, the Region will continue to focus future funding on breaking down barriers identified by the CDE.

Region Hovedstaden

KØBENHAVNS KOMMUNE

Hitachi Consulting

COPENHAGEN SOLUTIONS LAB

Vækstforum Hovedstaden

Smart City Insights

WANT TO KNOW MORE?

Please contact:

Marius Sylverstersen:

ms@cphsolutionslab.dk

or

Henrik Aagaard Johanson:

henrik.johanson@regionh.dk