BIG DATA ANALYTICS AND DECISION MAKING

Exploring the Possibilities of Big Data Standards to enhance the Management of Urban Development and Construction Projects in the Built Environment

Maren I. Skinner

- THIS PAGE WAS INTENTIONALLY LEFT BLANK -

Colophon

BIG DATA ANALYTICS AND DECISION MAKING
Exploring the Possibilities of Big Data Standards to enhance the Management of Urban Development and Construction Projects in the Built Environment

MAREN IONA SKINNER
Student Number: 4625625
Email: maren.skinner@gmail.com
Phone: +49 (0) 174 320 6629
Address: Warmoezierstraat, 9A, 2613 VG, Delft

DELFT UNIVERSITY OF TECHNOLOGY
Faculty: Architecture, Urbanism and Building Sciences
Master Track: Management in the Built Environment
Graduation Lab: BOLD Cities
Lab Coordinator: Dr. Ir. Alexander Koutamanis

Address: Julianalaan 134, 2628 BL, Delft
                     PO box 5043, 2600 GA, Delft
Phone: +31 15 2789111
Website: www.bk.tudelft.nl

MENTORS
First Mentor: Prof. dr. E.M. van Bueren
                     Management in the Built Environment
                     Urban Development Management
Second Mentor: Dr. A. Mauri
                     Management in the Built Environment
                     Urban Development Management
Third Mentor: Dr. A. Ersoy
                     Management in the Built Environment
                     Urban Development Management

- July 2018 -

- THIS PAGE WAS INTENTIONALLY LEFT BLANK -

- THIS PAGE WAS INTENTIONALLY LEFT BLANK -

Preface

Belonging to the generation that grew up in the midst of the transformation from *analogue* to *digital*, I had the privilege of experiencing the best from both sides: as a young child, I was able to play outdoors in the quiet suburban streets of my hometown, preparing a 'salad' from freshly picked daisies and building a treehouse in my family's garden. All this without the distraction of gaming apps or social media.

During my school years, I was introduced to computers and mobile phones at an age when it was easy to learn about new technologies and they really made life easier.

By the time I got to university, I was so thankful for the progress of digitization! Just imagining I would have to draft all my architecture drawings by hand was a nightmare.

Now, that we are far into the digital age and it is almost impossible to imagine an analogue life, even more advanced, sometimes even extreme, concepts are on the rise. Looking back on a wonderful childhood without all these digital advancements, I ask myself at times how much further the transition to *digital* will go.

My cautious stance towards digitization, smart technologies and big data analytics together with my strong interest for placemaking, building in existing contexts and contemporary urban challenges made the graduation lab *BOLD Cities* the perfect choice for my thesis project.

Inspired by Richard Florida's newest publication *'The New Urban Crisis'*, I dove deeper into the contemporary challenges cities are facing today. The most prominent of these challenges being segregation and inequality . Despite the fact that cities are aware and tackling these issues in various ways they are still prevailing.

Intrigued as to the reasons why it has so far not been possible to resolve such issues, I discovered a gap in the contemporary approach toward managing urban challenges. This gap, simply put, depicts an information shortage or a lack of data, that could potentially be filled through new means of information collection, namely big data analytics.

I set out to explore the link between big data and urban development management. The aim was to identify possibilities of utilizing big data standards for the purpose of enhancing decision making processes of urban development and building construction projects, ultimately to tackle issues of segregation and inequality in cities.

At this point, I would like to take a moment to thank the people, that have helped and supported me while researching and writing this thesis over the last nine months. My first round of 'thank yous' goes to my mentors from the UDM department Ellen van Bueren, Andrea Mauri and Aksel Ersoy, who were always positive throughout my research process, guided me with professional feedback and boosted my confidence when I felt unsure. Apart from that, I want to thank the practitioners at Rochdale and Ymere, for the time they spent on contributing to my survey and during the discussion sessions. These meetings gave me new perspectives and without this feedback my work would have remained one big guess. Thus, my gratitude also extends to Vincent Gruis, who made the introduction to Ymere.

Not only the professional encouragement and advice I received, has brought me this far. Thank you to my family and friends for always having my back, the motivational prep-talks and diverting food comas due to countless dinners, breakfasts or ice creams savoured in great company. I couldn't have done it without you.

Maren Iona Skinner
Delft, July 2018

Management Summary

Global trends such as digitization and the Internet of Things have influenced cities and the lives of residents deeply in recent years. In 2020 there will be an entire generation - *Generation C* - that has grown up connected. The influence of these trends is not only perceptible in the lives of individuals, but also entire governments are adopting so-called smart city concepts, which aim at raising the living standards of urban areas to accommodate the fundamental needs of the growing number of inhabitants. Such a smart city relies on the infrastructure provided by a digital city, which is created through a network of data analytics. This network is established through the collection of data, which has recently lead to the accumulation of large data amounts, depicting high resolution and real-time information.

Apart from these new developments of unexplored big data opportunities, there are long-standing urban challenges, such as inequality and segregation, governments and individuals are facing. In European cities and metropolitan regions, the prevailing inequality and residential segregation has developed into so-called 'poverty concentration areas', which are deprived districts where mainly immigrants, lower income groups and the less fortunate cluster.

Governments' responses to the prevailing issues of inequality and segregation have become the initiation of redevelopment projects, with the aim of dissolving areas of poverty concentration and creating an urban social mix and economically strengthened neighborhoods. More often than not these interventions start from the physical component. Through demolition and new build, renovation or densification residential units are developed, fit to accommodate middle income groups. It is expected that these state-led projects, through an attractive housing stock, will attract the mid-income group. Their settling in the district will ultimately create the anticipated urban social mix and economically strengthen the area.
This approach, however, leads to negative consequences.

A lack of information vital to the success of such governmental redevelopment interventions has been identified. It seems that the contemporary approach neglects to take sufficient information on the complex social dynamics of targeted neighbourhoods into account. This consequently leads to the displacement of resident groups and results in exclusive developments - the contrary of what the anticipated aim of the intervention was. The initially tackled urban challenges of inequality and segregation are merely relocated and while on paper, a neighborhood might seem more socially mixed or economically strong after such an intervention, on a more closer look it turns out that it is more divided than before, as individuals from different income groups barely interact, and economic benefits only reach some - most of the times the ones that came to the area with more already. Through the exclusiveness of these developments the residential segregation is furthered, even more so through displacement, which has the negative effect that social support structures, on which the less fortunate depend rigorously, are uprooted. All in all, these negative consequences of state-led interventions should be mitigated. By addressing the above concerns already in the planning and decision making phase of urban development or building construction projects, the issues could preventively be tackled.

As the possibility for negative consequences to arise is attributed to a lack of data, traditional information collection methods are examined carefully.
From literature it is identified, that traditional information indicators circle around data on the physical environment and the people within. The collected information does not extend beyond socio-economic indicators and thus, does not include data on intrinsic local structures, which actually define the area and the social relationships found within. Not only that, but the available information originates from census and registry data, which is likely to be several months, if not years old and is usually

distributed on a postcode level. Hence, traditional information collection methods are not only prone to depict outdated and generic data, but also oversee entire aspects of essential information.

Exploring new possibilities of information collection, to ensure high resolution and real-time data, as well as information on intrinsic social structures of areas, can be delivered to policy makers and planners during planning and decision making phases, is key to mitigating the negative consequences of contemporary redevelopment interventions. New information collection methods could support more comprehensive studies of urban social dynamics enabling better management of urban developments and building construction projects, to arrive at more successful and inclusive (re)developments.

Here, digitization or rather *Big Data* offers new possibilities. To understand the concept of *Big Data* better, it is oftentimes broken down into the multiple characteristics, the so-called Vs, of data management: *Volume, Velocity, Variety, Variability and Value*.

This lineup of *Big Data* characteristics demonstrates the range of potential opportunities. Overall, big data analytics has the advantage of making a considerable amount of information available without the need for information extraction through traditional information collection methods, it enables real-time access to information and allows examination and analysis of data on a higher resolution. Big data analytics describe an unconventional new way, of information collection from social media sites and online platforms. The advantage of these data sources in particular is also that information can be linked to a place, so-called geo-locational data. This makes it possible to derive information embedded in a local context.

Thus, the main research question was formulated as follows:

How can big data standards be utilized to manage city dynamics and enhance planning and decision making processes of urban development and construction projects in the built environment to mitigate the negative outcomes of such state-led interventions?

Based on an extensive literature review, assumptions were formulated with regard to the information that can potentially be collected through the big data sources insidearibnb.com and social media, particularily Twitter, Flickr and Foursquare.

A critical review of these assumptions then lead to the formulation of ten big data information collection indicators. While four of these add information directly to the traditional aspects of properties and the people, the remaining six big data indicators focus on collecting information on the spatial level, thus reflecting data on intrinsic social structures of neighborhoods and potentially filling the gap in traditional information collection.

As the information collection on traditional indicators and big data indicators is interconnected, and only in combination a truly holistic neighborhood profile is depicted, the contemporary approach is complemented with the big data information indicators, arriving at an integrated framework for information collection.

The performance of big data indicators, meaning the ability of big data standards to extract valid and valuable information, was illustrated through a case study analysis and the input of practitioners. This brought forth interesting findings.

While overall, the performance of big data indicators was successful, the outcomes of the illustration are bound by several conditions.

The most prominent being the dependency of the integrated framework's performance on a data rich environment. An area in which social media is barely used or Airbnb locations are scarce will not be able to produce a sufficient data output. Secondly, Big Data is not able to produce valid information outputs on its own. This is also why an integrated framework between traditional and big data indicators was formed. But also within the individual big data standards for information collection

correlation between various data sources is necessary to ensure the validity of the retrieved information.

Moreover, it was found that the choice of data sources also impacted the outcome. Insideairbnb.com utilizes data from Airbnb.com, which is a platform offering short-term rentals and booking services. In literature the speculation has been voiced, that Airbnb is mainly frequented by mid-income Westerners, meaning any information derived from this sources merely depicts data with regard to that specific part of the population. A similar circumstance can be found with social media. The general opinion is that social media is inaccessible. This medium is mostly used by different social groups under the age of 50. However, even between the various social media platforms there are differences. Twitter is used throughout all age groups, while Flickr and Foursquare are more heavily embraced by individuals aged 35 and under. This of course leads to the underrepresentation of the population group born in the 80s and before.

Apart from that, some of the information that is assumably collectable through big data standards is just not retrievable in the intended way. In other words, some anticipated theoretical results do not uphold in practice.

These identified conditions reflect the immaturity of big data analytics. While the three main advantages of *Big Data*, namely gathering data on a high resolution, in real-time and extending the information collection to social and spatial aspects of the area, could be confirmed in the context of this research, there are many more matters to be resolved, before the integrated framework or similar can be put to practice.

The aim of this research has been to investigate, whether big data standards hold opportunities to enhance the management of urban development and building construction projects and thus mitigate the negative consequences. The integrated framework gives a positive inclination towards this question and is a first step towards a new approach of information collection, that is almost ahead of time - in another 20 years, when *Generation C* has their own children and the beforehand mentioned global trends are progressing further, more and more parts of the population will be connected and thus, represented in data sets, such as the ones proposed above.

As such, this research is of great informative value for both theory and practice of the arising digital age, as it holds the knowledge base on which future research in this field can build upon.

Reader's Guide

Chapter 1-3 [Research Proposal]

Introduction to the Topic, Research Questions, Societal and Scientific Background of the Research Issue

Chapter 4-5 [Research Procedure]

The Research Methods, Data Handling, Setting up a Framework and Case Study Analysis

Chapter 6-7 [Research Product]

Illustration and the Performance of the Integrated Framework, Practical Implications

Chapter 8-9 [Research Results]

Conclusions, Relevance, Recommendations and Reflections

Content

List of Figures

List of Tables

- THIS PAGE WAS INTENTIONALLY LEFT BLANK -

1. Introduction

Global trends such as digitization and the Internet of Things have influenced cities and the lives of residents deeply in recent years (Sassen, 2005). In 2020 there will be an entire generation - *Generation C* - that has grown up connected. The influence of these trends is not only perceptible in the lives of individuals, but also entire governments are adopting so-called smart city concepts, which aim at raising the living standards of urban areas to accommodate the fundamental needs of the rising number of inhabitants and improve the overall performance of health, transportation, energy, education and water services within cities (Nuaimi et al., 2015). Such a smart city relies on the infrastructure provided through a digital city, which is created through a network of data analytics, forming an information model that covers the entire city. Through this model the city's scattered information of various types is organized based on geographical locations (Li et al., 2015).

Big data analytics begins with the collection of data, which has recently lead to the accumulation of rather large data amounts from networks of sensor data from which a digital platform deals with *'the massive data storage, computation, analysis, and decision-making process, and conducts automated control based on the results of those analyses and decisions'* (Li et al., 2015).

Besides the advantages of collecting and analysing complex, detailed and real-time information, the drawbacks of big data need to be regarded. Particularly, a smart city built on big data needs to take into account big data challenges with regard to the availability of big data tools, real-time analytics, certainty, representation, cost and accessibility (Nuaimi et al., 2015), but also such issues as cyber security and privacy concerns.

The MBE research in BOLD Cities centralizes around different perspectives of big data, such as architectural and planning information, smart cities information technologies, urban area development and governance. A main focus is discovering the potential of big data for cities. In environments like living labs and different societal concerns, stakeholder participation plays a prominent role.

Up until now, research products include BOLD-based analyses that combine use and building data, evaluations of existing Dutch cases and apps that bring BOLD to the citizens.

Apart from these new developments and unexplored big data opportunities, there are long-standing challenges and the question of who the city belongs to, that call for attention. The talk is of an 'urban glamour zone' versus an 'urban war zone' resulting from opposing groups making claims to the ownership status of a city (Sassen, 2010). According to Richard Florida, these rivaling zones are reflected in the growing inequality, especially in large cities, which manifest themselves in the growth of poverty and economic disadvantage (Florida, 2017). This growth of poverty and economic disadvantage is revealed in the emergence of an 'urban underclass' in an 'urban war zone'. However, over 20 years ago, Sako Musterd already pointed out the lack of opportunity for people who are socially excluded (Musterd, 1994). He titled the growing separation of higher sector jobs from lower sector jobs as "social polarization hypothesis", which can also be described as socio-spatial polarization.

It can be concluded that, social inequality, meaning the disadvantaged poor, also leads to socio-spatial inequality or polarization. This socio-spatial inequality is part of the phenomenon of socio-economic segregation, which describes the residential segregation of population groups based on occupation, income and education (Musterd et al., 2017).

Manfred Kühn (2014) portrays this polarization or segregation circumstance as 'peripheralization', which is the process of producing peripheries through social relations and their effect on spatial matters (Kühn, 2014).

What Kühn (2014) refers to as peripheries in his research are mainly rural or former industrial areas that suffer under the rapid growth of centers. This can be seen in North American metropolises, where

cities are often monocentric (57,7% of metropolises in 2010) with a flourishing core and underprivileged surroundings (Florida, 2017; Arribas-Bel & Sanz-Gracia, 2014, Hall, 1997).

European cities, however, are more often polycentric. This has to do with the European Spatial Development Perspective (ESDP), which advocated greater polycentricity in the European urban system in 1999 (Hall, 2006). However, this does not mean that polycentric European cities do not face the urban challenge of deprived areas. It is just not obvious that these deprived areas are found in the periphery of a metropolis. More often, underprivileged districts in European cities and metropolitan regions are scattered over the area, alongside thriving sections. To illustrate, Turin, Amsterdam and Stockholm have been picked as example cities, having a comparable size and number of total inhabitants (Figure 1-3). Areas of highest unemployment rates (coloured brown) are located next to areas of lowest unemployment rates (coloured dark green) (University College London, n.d.).
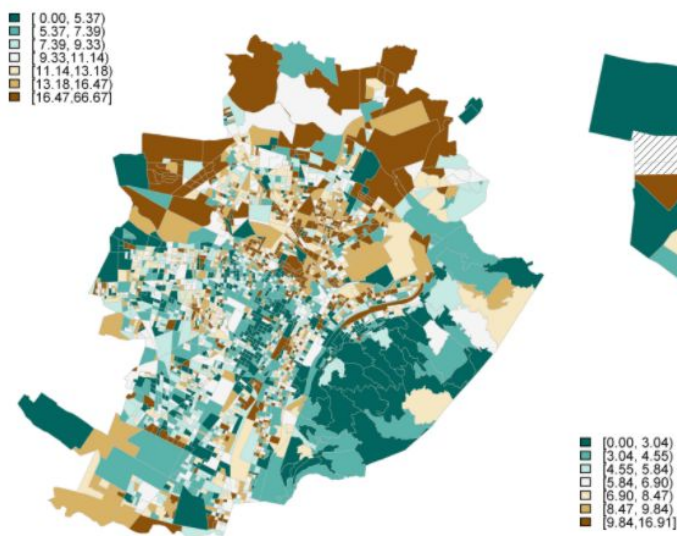
Figure 1 - Unemployment Turin (source: https://www.ucl.ac.uk/ineqcities/atlas/cities/turin/turin-sei)
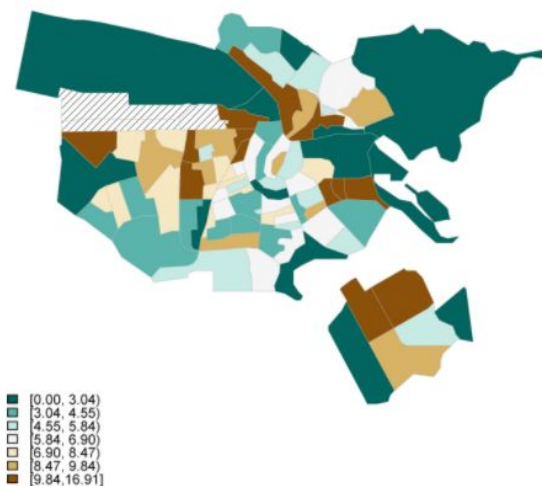
Figure 2 - Unemployment Amsterdam (source: https://www.ucl.ac.uk/ineqcities/atlas/cities/amsterdam/amsterdam-sei)

Figure 3 - Unemployment Stockholm

(source: https://www.ucl.ac.uk/ineqcities/atlas/cities/stockholm/stockholm-sei)

This study on unravelling the opportunities of big data standards in decision making to enhance the management of city dynamics, will be framed within the context of the urban renaissance in Amsterdam.

More often, the government's response to the prevailing issue of inequality has become the initiation of redevelopment projects, with the aim of creating an urban social mix and economically strengthening areas of poverty concentration. Information on the complex social urban dynamics of targeted neighbourhoods is essential to ensure the success of such governmental redevelopment interventions. Observing the outcome of so-called 'state-led gentrification projects' and through this illustration, exploring the advantages big data analytics hold to enhance the management of urban development and building construction projects, is the overall ambition. Apart from that it is investigated, whether big data can function as a vehicle to mobilize the stakeholders involved in such processes. Also, it is examined in what way big data analytics are of value in the urban development practice, whether they can empower practitioners and what the prerequisites are for such standards to be put to practice.

The main research question guiding this explorative study aims at determining *How big data standards can be utilized to manage city dynamics and enhance planning and decision making processes of urban development and construction projects in the built environment to mitigate the negative outcomes of such state-led interventions?*

 To support this central query, sub-research questions corresponding with this reports chapters, have been formulated, first focusing on the origin of inequality and the context of urban development in Amsterdam. *How does inequality prevail in cities today? How do cities like Amsterdam cope with inequality? What are the predominant negative consequences of contemporary urban development and building construction processes?* Followed by the inquiries on the role of big data*. What are the possibilities of big data with regard to challenges in the urban context? What are the drawbacks of big data analytics?* Questions regarding the procedure of big data possibilities are: *How can big data be utilized for the purpose of managing city dynamics? How can big data be utilized for the purpose of enhancing planning and decision making of urban development and building construction processes?* Finally, the results of an illustration of big data analytics within a context will shine light on the question *How big data standards capable of managing and enhancing decision making process in the context of urban development and building construction projects look like?* and consequently *How big data can be utilized to manage city dynamics and enhance planning and decision making of urban development and construction processes in the built environment*?

On top of this main research body, sits the certification of the analysed big data standards. *Are big data standards able to empower practitioners within planning and decision making processes?* and *What are the practical requirements to implement such standards in practice?*

The research begins by unraveling the origins of inequality, leading from the analysis of historic city dynamics to newer and contextualized developments with 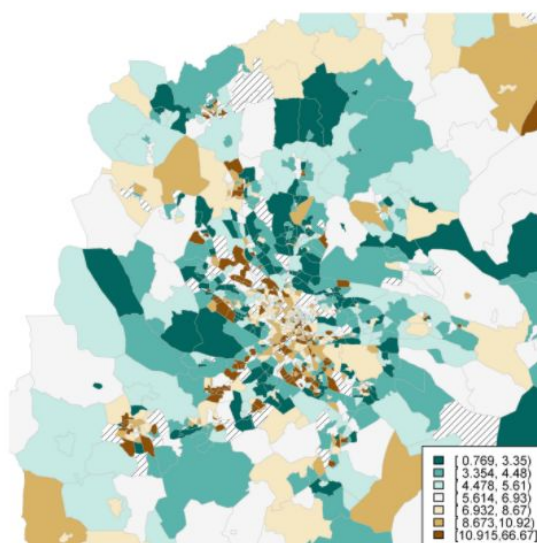regard to socio-economic segregation. The aim is to better understand social urban dynamics in their multifaceted ways and determine the disadvantages of the contemporary responses in form of state-led gentrification projects to prevailing inequalities.

This is followed by the introduction current big data standards. Their opportunities and limitations are thoroughly analysed and relevant data sources are introduced.

Next, the anticipated methods for this research are explained in detail, including the means and added value of each of the techniques.

Through the illustration of the introduced big data standards within the context of Amsterdam Nieuw West, the uncovered possibilities are substantiated.

Eventually, the outcome of this research in form of an integrated framework is introduced and it is reflected on the findings.

Additionally, it will be established whether the integrated framework depicts the basis for future research to develop tools able to support stakeholders and empower practitioners during decision making process of urban development and building construction projects.

2. Origin of Inequality

2.1. Urban Gentrification

Part of the long-standing challenges cities are confronted with is the fact that they are undoubtedly facing socio-economic segregation processes (Musterd et al., 2017; Kühn, 2014). But, what makes certain neighbourhoods of cities decline, while others experience growth? What are the drivers of polarization processes?

Kühn (2014) poses answers to this question and states theoretical approaches to explain these processes, starting from economic polarisation and social inequality (Kühn, 2014). He describes, that, in a knowledge-based economy, centralization occurs through attracting people. The economic productivity of such economies is based on attracting highly educated professionals, which start inhabiting the area, driving housing prices up (Cucca, 2012). This leaves little room for the lower income class and almost consequently leads to social inequality. The service worker might be attracted by the economy of a place, however he cannot afford to locate there, which leads to socio-spatial inequalities. Places attracting educated professionals grow and thrive, while places inhabited by the lower educated become socially and spatially excluded (Kühn, 2014). What this boils down to, is the fact that polarization processes are driven by two things: what a certain place has to offer as well as the quality of its supply.

The process which Kühn (2014) describes can be summarised in the term *gentrification*. This terminology originates from the critical observation by Ruth Glass (1964), who conveyed the displacement of the working class through the arrival of 'a new gentry' in many London neighborhoods, resulting in inequality. While the arrival and departure of different social groups to and from the city is part of the natural urban dynamics, segregation based on income, education or employment is not. Why in this case the 'new gentry' decided to settle in certain areas was dependent on many different factors, making this phenomenon quite a complex one to grasp. Today, this complexity is reflected in the multiplicity of definitions that are offered to explain the term gentrification. According to Galster & Peacock (1985), it needs to be acknowledged that gentrification is driven both by characteristics of people and of property, similar to the duality driving centralisation, referred to by Kühn (2014).

In 1996, Neil Smith, one of the leading gentrification theorists, known for his rent (gap) theory, defined gentrification as, the 'symbiotic change in social class and physical housing stock' (Smith, 1996). Again, referring to people and property and leading to the conclusion that important factors or indicators of gentrification involve a combination of income and rent data (Smith, 1996).

A more recent definition, which covers a wider variety of aspects of gentrification compared to Smith (1996), without being too restrictive, has been put down by Fenton (2016):

> "[...] gentrification is understood as changes in the composition and character of urban localities in favour of newer residents with more resources and at the expense of existing residents with fewer resources;"

Through which criteria these changes in the composition of an urban area are determined and which factors are considered to be indicative of gentrification can differ from country to country, but also from city to city, depending on the historical, political, cultural and social circumstances. This is confirmed in the publication by Galster & Peacock (1985), where it is stated that both the definition of gentrification matters immensely, as well as the indicators chosen, as they can influence the research outcome heavily (Glaster & Peacock, 1985). Therefore, it is necessary for this research to define not only the outline of gentrification through a suitable definition, but also identify relevant indicators of gentrification. The above stated definition by Fenton (2016) is considered adequate for this research, as it touches upon all relevant aspects of gentrification. It includes both people and property, but adds

the layer of area to the definition, while being open ended regarding the exact characteristics. This is assumed to be of value later on when gentrification indicators in the built environment are identified and set within an urban development context .

As outlined earlier, what most definitions have in common, is the fact that gentrification brings with it a substantial change to the composition of an urban area. However, opinions differ regarding the effect these changes have on the neighbourhood and the people living within, resulting in major disputes regarding the benefits of gentrification.

2.2. A Paradox

Ever since the urban phenomenon of gentrification has been defined, opposing points of view have been formulated. On the one hand, it is claimed that attracting the affluent and well-educated is good for the city, as they are more able to finance social housing through taxes and raise the overall political and economic position of a city (Byrne, 2003; Atkins, 2002). On the other hand, the assumption that the benefits of gentrification would also reach the lower classes, has been firmly criticized. Moreover, gentrification is believed to actually encourage displacement, segregation and social polarisation, rather than easing these occurrences (Lees, 2008; Uitermark, 2003).
However, according to the Byrne (2003), it is wrongful to describe gentrification as the 'culprit' in the process of displacement, as neighbourhoods have always been dynamic. Naturally, newcomers and residents have different viewpoints, but this does not only create problems.
Looking more closely at the claimed benefits of gentrification, they can be classified within four categories: *economic, political, social and physical.* Economic benefits of gentrification include new employment opportunities, a more accessible housing market through easier mortgage lending and improved shopping and economic activities ((Byrne, 2003; Atkins, 2002). Political benefits are the creation of larger political entities for the disadvantaged to be heard and the strengthening of the negotiation power of disadvantaged through the city's population growth (Byrne, 2003). Social benefits focus on the reduced isolation of social groups, reduction in crime and better equipped educational institutions (Byrne, 2003; Atkins, 2002). Finally, the physical benefits of gentrification relate to the revitalization of the physical stock of the city (Atkins, 2002).

While these benefits, resulting from empirical work, make a relevant point, it cannot be denied that people have experienced displacement, directly or indirectly, through the dynamics which a neighbourhood or city is subject to.
In other words, there is no sufficiently researched proof, that the assumed advantages or disadvantages are coming to pass. The literary evidence is torn between the opinion that gentrification is beneficial for the city and the opinion that gentrification is harmful to individuals. However, since there seems to be even the slightest belief in the harmfulness of gentrification to citizens, there is enough evidence to research this further, especially, since gentrification is an arising policy concept in North America and Europe. The doubts regarding the problems and opportunities gentrification creates and its effects on the city and its citizens are not inhibiting the implementation. The policies build on the assumption that gentrification will increase the social mix, encourage social mixing and thus increase the social capital and social cohesion of inner city communities (van Gent, 2013; Lees, 2008; Uitermark, 2003), which is why 'state-led gentrification projects' are an emerging answer to these long-standing issues. However, as stated before, due to the lack of fundamentally researched proof, these anticipated positive effects are questionable. Why the policies and such projects are nonetheless implemented is elaborated in the following.

2.2.1 Gentrification as a Policy

For some time, populations all over North America and Europe have been confronted with policy terms such as urban renaissance, urban revitalisation, urban regeneration or urban sustainability (Lees, 2008). What lay-people have difficulty discovering is the fact, that these terms cloud the actual policy aim, namely a cleansing agenda, enabled through gentrification (Cameron, 2003). As stated before, governments have come to believe, that urban issues, such as segregation and inequality, can be solved through social mixing. Criticism to this policy standpoint is deflected through not utilizing the term 'gentrification' directly and also by developing the aspect of social mixing into a moral duty (Lees, 2008). Gentrification is turned into a 'positive public tool' (Cameron, 2003), which is assumed to bring about a new urban renaissance. These social mixing policies tend to utilize the built environment to reach the aim of reducing inequality in the city. A positive example is the revitalization of downtown Los Angeles. Through investing in the building stock and the redevelopment of housing, the municipality was able to attract desired tenants, facilities and cultural institutions and turn the once dangerous downtown area into a safe and walkable neighbourhood.

Even though the Netherlands is dealing with comparatively low shares of spatial segregation, social segregation and social mixing are on the government's agenda. The policy housing redifferentiation foresees the expansion of expensive dwellings within low income areas, through the demolition of social housing units. Generally, such policies can also have positive effects, as has been mentioned and illustrated above. However, the motivation behind this particular housing redifferentiation policy does not have the interest of the poor individual at heart, rather the aim of strengthening the economic position of Dutch cities (Lees, 2008; Uitermark, 2003). Also, research by Uitermark et al (2007) has made it clear, that social cohesion is not increased by the arrival of the middle class into a low income neighbourhood. Therefore, the citiques mentioned earlier are also valid for review within the Dutch context.

2.2.2. Urban Renaissance in the Netherlands

With his 2003 research, Uitermark is trying to explain why Dutch policy makers are pursuing social mixing and where and when they do so. According to the official social mixing policy of the Netherlands, the social management of deprived urban areas, as part of the city-oriented national growth strategy, has become a political priority (Uitermark, 2003). As mentioned in the case above this policy foresees to reduce social housing through demolition and essentially replacing it with better quality housing to attract the affluent (van Gent, 2013; Lees, 2008; Uitermark, 2003).
The implementation of this policy, despite the lack of evidence on the benefits and the disadvantages of gentrification, can be explained through the interrelationship between the state and institutional actors.
A rescaling of responsibilities has been taking place, as a shift from government to governance. Whereby, the Dutch government has been particularly focused on the demands of institutional actors in deprived urban areas. The 'polder model', an institutional structure that has evolved during the 1990s, displayed the commitment of social groups to solve conflicts through cooperation and compromise. However, this seemingly harmonious but delicate system was under threat by multiple issues, first and foremost the difficult integration of ethnic minorities (Uitermark, 2003). Growing tension resulted from the continuing socioeconomic position of these groups and concerns regarding the feasibility of the polder model. These issues seemed to be threatening not only certain neighbourhoods, but the whole economic status of large Dutch cities. Which lead mayors of these cities to demand intervention by the state by strengthening the position of relatively poor cities and neighbourhoods. The focus of these initiatives was on the economic aspects and therefore, seemed to have conflicting consequences for the disadvantaged. Gentrification policies were implemented with the aim of economic optimisation. Tension between wealth creation and wealth distribution has

arisen. But the question remains, whether real economic values can actually be reached with such a strategy (Doucet et al., 2011).

What this historic and political recap, based on the research by Uitermark (2003), leads up to is the fact that the interrelations between national actors and neighbourhood managers now link these two levels most intricately. This kind of regulation on the neighbourhood scale is believed to compensate for shortfalls on the national scale, reversely making deprived places easier to manage through government intervention and simultaneously clearing the way for the city-oriented national growth strategy.

With this background information it is possible to better understand policy-led gentrification processes in the context of the Netherlands and how it was possible for this economic-oriented policy to emerge.

As the above literature review shows, arising policy-led gentrification processes are impacting natural urban social dynamics. On occasion, these gentrification policies have negative consequences for the city and its residents. What these negative consequences look like and how they have developed in Amsterdam is the topic of the following chapter.

2.3. State-led Gentrification Projects in Amsterdam

The capital of the Netherlands, Amsterdam, has been a highly valued target of such policy-led processes, such as the restructuring of the housing market, transformation of the built environment and essentially gentrification (van Gent, 2013). After undergoing so called first and second phases of gentrification, Amsterdam has been following a rather aggressive, state-led pursuit of gentrification since the 1990s. Enabled through privatization, decentralization and deregulation, the housing system changed, as did the roles of actors. Policies actively promoted homeownership, believing that a change in tenure composition would positively influence the social mix (van Gent, 2013). The municipality of Amsterdam demonstrated the mismatch between household income groups and housing stock, therefore justifying the state-led gentrification as a means to move lower middle income households to owner occupied dwellings and thereby freeing up social-rental dwellings for the low income classes (Municipality of Amsterdam, 2008). However, this venture has brought about a turn in the responsibilities of serving housing needs. Housing associations, traditionally set with the task of providing basic housing, are now answering to more market-oriented needs, providing less social-rented dwellings in turn for developing owner occupied housing. Generally, more market oriented developments are taking place, while less funding is mobilised for social-rental housing and social regeneration efforts (van Gent, 2013).

Continuing with the case of Amsterdam, there is more behind the current regeneration policies than just strengthening the economic position of the city (van Gent, 2013). The municipality of Amsterdam is also driven by the aim of providing the right amount of housing for each household type. Through which, issues such as spatial isolation, the occurrence of physically deprived neighbourhoods with high crime rates, next to no employment opportunities and lacking facilities, shall be mitigated. However, what seems to be disregarded are the means by which these aims, regarding economic, political, social and physical goals, are being enforced.

The built environment seems to be utilized for the municipality's purpose without regarding the local context. The policies in question are being imposed by national and municipal governments and focus on the nationwide and city scale. However, the inequalities prevailing in cities cannot be tackled without taking also the complex local structures into account. Despite all intentions, these policies more often result in exclusive developments and displacement (Lees, 2008; Uitermark, 2003; Doucet, 2011) This is why the possibilities of big data analytics in decision making of urban development and building construction projects are reviewed, to determine, whether through these standards it is possible to collect more detailed information on city dynamics and thus, secure nationwide city growth as well as facilitate a more equal approach to neighbourhood management.

By analysing the indicators of gentrification derived from literature in the following, it can be determined in which areas the decision making and planning process of urban development projects could be enhanced through big data standards.

2.4. Indicators of Gentrification

The concern raised above is reflected in the traditional approach toward gentrification research. Contemporary studies on gentrification in the urban context have been examined throughout this chapter and by collecting the indicators, meaning the variables that are typically analysed during a gentrification study, a gap has been identified.

		U.S.	UK	The Netherlands
PEOPLE	**INCOME**			
	Median Income	x	x	x
	High Income	x		
	Low Income	x		
	EDUCATION			
	Education Level	x	x	x
	Education Status			x
	EMPLOYMENT			
	Employment Type	x	x	
	Employment Status		x	x
	DEMOGRAPHIC STRUCTURE			
	Elderly	x		x
	Black	x		
	Origin	x		x
	Population Change	x	x	
	Population Growth			x
	Marital Status			x
PROPERTY	**HOUSING**			
	Tenure	x	x	x
	Age	x	x	x
	Unit	x		
	Contract Rent	x		x
	Average House Price			x
	Occupancy Rate		x	
	LOCATION			
	CBD	x		
	University	x		
	HIST	x		x
	Park	x		
	TURNOVERRATE			
	Private Property	x		
AREA	**VACANCY**			
	Abandoned Structures	x		
	HEALTH			
	Distribution of Practicioneers		x	

Table 1 - Traditional Indicators Worldwide

Table 1 depicts the indicators traditionally collected in gentrification research in three countries, the U.S., UK and the Netherlands (Galster & Peacock, 1985; Smith, 1996; Wheeler, et al., 2005; Fenton, 2016; Divercities, 2017; Salvati, 2017). The literature on gentrification research within these territories is especially rich, which is why they are dominant in this country comparison.
The identified indicators have been grouped within the fields 'people', 'property' and 'area', which reflect the three themes essential to the definition of gentrification as stated by Fenton (2016) earlier. Indicators regarding topics of income, education, employment and demographics belong to the field 'people'. Indicators concerning housing, location or turnover rates are part of the field 'property'. Finally, indicators regarding vacancy or health are placed in the field 'area'. It is important to notice that, overall, the focus of gentrification researchers lies on indicators within the fields of 'people' and 'property', as 24 out of 26 collected indicators fall within this category, while the field 'area' seems to be barely examined (only 2 of the 26 most commonly applied indicators). This is an alarming discovery, as the 'area' has been determined to be an essential aspect in processes of gentrification.

Based on this international comparison of traditional indicators, the variables relevant for the Netherlands, and specifically, the issue of this research, are combined in table 2. The suggestion from the results of the country comparison, that indicators from the field 'area' are disregarded, is confirmed for the Netherlands.

The Netherlands	
PEOPLE	Median Income
	Education Level
	Education Status
	Employment Status
	Elderly
	Origin
	Population Growth
	Marital Status
PROPERTY	Tenure
	Age
	Contract Rent
	Average House Price
	HIST

Table 2 - Traditional Indicators in the Netherlands

Before continuing with the meaning and consequences of this identified gap for urban development and construction projects, the traditional method of information collection of each variable is briefly introduced in the following. This builds up the knowledge to better evaluate the opportunities and challenges of information collection through big data standards.

Indicator Field 'People'
The information with regard to variables belonging to the field 'people' are mainly researched through official census data. In the Netherlands this is done through the organisation CBS Statistic Netherlands, an autonomous administrative body, performing public service tasks, financed by the state budget (CBS, n.d., A). The data published by CBS has a direct impact on Dutch citizens' lives. CBS are responsible for compiling (official) national statistics, as well as European (community) statistics.

Until 1971, the Dutch population census was performed by sending individual questionnaires to each household across the country. From 1981 onwards, the system was renewed and the population and housing census was based on registry data. Registry data combines data from municipalities, the tax department, employment organisations, the land registry, chamber of commerce and the Dutch central bank, and today multiple other data sources as well. In 2011 CBS had 600 terabytes of data available. The idea is to create a 'virtual census', not by adding new data but by connecting existing data sources and combining them with registry data (CBS, n.d., B).

Even though the amounts of data have vastly expanded and more detailed data is accessible to CBS, such a census is performed only once every ten years.

The introduction of variables below, has the aim to highlight, first of all the data traditionally considered valuable in gentrification research and secondly, the sources from which they are currently retrieved.

Median Income
the average monthly income of a population group within a specified area. This data can be anonymously retrieved from employment organisations, the tax department and the Dutch central bank.

Education Level
the years spent on school and higher level education. The data can be retrieved from employment organisations and municipalities.

Education Status
the current state of an individual's enrollment within an educational institution. This data can be retrieved from the municipal registry.

Employment Status
the current state of an individual's employment. The data can be derived from the tax department and the municipality.

Elderly
the age and health condition of an individual. This data is accessible through the municipality.

Origin
the ancestry, specifically the country of origin of an individual. This is retrievable through the municipal registry.

Population growth
a population's development over time. This data is retrievable through the municipality.

Marital Status
the family position of an individual. The data can be retrieved from the tax department and municipal registry.

Indicator Field 'Property'
Data on variables from the 'property' field can partly be collected from similar data sources as the information required on 'people'. Using information from registry and other data sources offers the same possibilities and limitations as described above. On the one hand, the pool of data has been enlarged allowing for more detailed information to be extracted, on the other hand, data is again sensitive to being outdated as it is compiled only every 10 years.
See the variables below, for an individual overview of traditional data and their sources.

Tenure
the tenureship status of a property. is it rented out, owner occupied, vacant etc. This data can be retrieved from the municipal registry and the land registry.

Age
the years that have passed since the original construction or since the most recent renovation of a property. The data can be found in the land registry.

Contract Rent
the rental amount to be paid by the tenant on a monthly, quarterly or yearly basis. This data can be retrieved from statistical real estate reports compiled by real estate firms, such as Savills, CBRE or Colliers or through housing corporations registry.

Average House Price
the median amount for which a house is sold within an area. The information is accessible through statistical real estate reports or through the municipal registry.

HIST
the proximity to historic sites. This data can be retrieved through the municipality and the land registry.

This brief examination of the individual indicators, shows how a larger volume of data has become available in recent years, reducing a cost issue, but that the challenge of retrieving up-to-date sets of information remains.
Moreover, this analysis confirms the statement formulated in the chapter above. The described gap of traditional information collection failing to take into account variables from the *area* category, such as social relationships and complex local structures, is confirmed. The focus lies on socioeconomic factors within the categories of *people* and *property*, collected from extensive surveys and census data. If this is the only information taken into account in planning and decision making process of urban development and building construction projects, the outcomes of 'state-led gentrification projects' in Amsterdam thus unsurprisingly result in unequal and exclusive developments. However, by complementing the above indicators with information on intrinsic local networks of urban areas, a more comprehensive study of urban social dynamics could be conducted, enabling better management of urban developments and construction projects, as well as enhancing their outcome, to arrive at more successful and inclusive (re)developments.

3. Management of Inequality

3.1. Big Data Standards

Studies on the question of gentrification, both the ones taking a positive and a negative stance, are largely facing limitations and exhibit research gaps. Most research studies, such as Atkins (2002) Byrne (2003), Lees (2008) or Uitermark (2002), just to name a few, focus on reviewing benefits and opportunities based on empirical evidence. The conclusions and results are hypothesized and conjectured, but hardly evidence based. Also, confirmation from case studies, or case studies other than the obvious candidates London and New York, is rarely integrated into the research.

Moreover, the topic of gentrification within empirical social science is strongly being restricted by the quantity of data available. Simultaneously, the fact that the information is generally being delivered in data collections, meaning in groupings such as 'city', 'occupation' or 'age', makes it nearly impossible to discover the distinct variations of individual social life (Rabari & Storper, 2013).

Not only the lack of detailed data and the distant level on which neighbourhood analysis takes place through gentrification studies, are troublesome. Also, the delay with which required information is released, due to the fact that data collection and evaluation processes are time consuming, results in the accuracy of the information being rather questionable, leading to a distorted picture of urban dynamics.

Here, big data analytics offer new possibilities. To understand the concept of big data better, it is oftentimes broken down into the multiple characteristics, the so-called Vs, of data management. According to Nuaimi et al. (2015) the five main Vs are *Volume, Velocity, Variety, Variability and Value*. *Volume* refers to the large amounts of data that are being created every day. *Velocity* concerns the speed at which data is created, saved, analyzed and processed every day, making real-time information available. *Variety* means the different types of data that are being created, as a lot of data is unstructured and cannot be easily classified. *Variability* refers to how the organization, significance and meaning of data changes continually. *Value* concerns the sum of these characteristics and how they possibly offer an advantage if based on good big data collection, management and analysis.

On occasion, literature offers additional Vs to describe more aspects of big data, such as *Volatility*, which considers the policy of retention for structured data implementation from various sources. Or *Validity* that deals with correctness, accuracy and validation of the data. Last but not least, *Veracity*, which means the accuracy and truthfulness of collected data and the significance of results generated from that data (Nuaimi et al., 2015).

This lineup of big data characteristics demonstrates the range of potential opportunities, which seems endless. The seeming only boundary being the available technologies and tools.

These potential opportunities are illustrated through a research article by Glaeser et al. (2015). The authors state, that big data is essentially helpful to measuring the physical city in general (Glaeser et al., 2015). This hypothesis is illustrated through various examples. "Digital exhaust", the trail of data left behind through individual's daily online activities, is a valuable but underutilized data source for monitoring urban issues, say the authors. Also, Rabari & Storper (2013) encourage the usage of open data to help measure (social) processes in cities. The authors use the term "digital skin", which visualises data as a sort of layer on top of the physical city. Both papers are genuinely positive regarding the advantage big data sources have for managing and enhancing complex developments in cities.

Despite these numerous advantages there are also limitations to current big data standards that require attention. It is believed, that by fully understanding the capabilities as well as the limitations, the many advantages can be better exploited, which is why big data limitations are introduced in the following.

3.1.1. Drawbacks of Current Big Data Standards

The predominant limitations of big data standards have been mentioned earlier. They concern the availability of big data tools, real-time analytics, certainty, representation, cost, accessibility, cyber security and privacy concerns. These can also be connected to the downsides of the Vs of data management. Concerns regarding the originality and representation are reflected in the sheer *Volume* of data becoming available, as well as in *Variety*, *Validity* and *Veracity*. The issue of real-time analytics or privacy are represented within *Velocity.* The availability of tools, accessibility and cyber security are displayed within *Variability* and *Volatility*. Cost concerns and the general matter of significance are depicted within *Value*.

In literature, more practical drawbacks are pointed out. However, they can almost always be linked back to the Vs of big data analytics. To illustrate, these the drawbacks from literature are introduced below and underlined with a practical example.

Correlation and Causality

According to Glaeser et al. (2015) and Rabari & Storper (2013), it is important to acknowledge that data, combined from different sources identifies only correlations and not causality (Glaeser, 2015; Rabari & Storper, 2013). The following example shall illustrate:

Assumably, the Twitter use in a neighbourhood has dropped sharply, since the launching of twitter in 2006 until now. Simultaneously, in the same neighbourhood, during the same observation period, the education level has fallen rapidly. Due to the prominent decrease in both at the same time and in the same place, one could say the events are correlated. However, identifying this correlation itself does not justify the interpretation that the usage of twitter has anything to do with the education level in a neighbourhood.

Small-number Phenomena

Another drawback, is the lack of big data being unable to identify small-number phenomena. This means, while it is possible through a big data analysis to collect specific features from a pool of data, which combined could identify a certain event, does not result in this event actually taking place. An example for this could be, the factors indicating the decline of a neighbourhood. Through big data analysis it would be possible to identify neighbourhoods with such factors and predicting decline before it takes place. However, due to external circumstances, not all neighbourhoods presenting such factors will actually experience a decline.

Echo-chamber Effect

The third drawback mentioned by Rabari & Storper (2013) is the so-called "echo-chamber effect". What the effect basically describes is the wrongful use of data by a collective that does not have the opportunity of consulting an alternative data source. This handicap is especially prone to occur in cases, where the information used for a data analysis is drawn from an internet source (open data). Oftentimes, the origin of the data is unknown, making the validity of the data questionable. Should it turn out, that this data contains false information, and no other information source is available to confirm otherwise, the false information will be carried into the data analysis undetected. Opportunities for "vicious cycles" might occur (Rabari & Storper, 2013), when the output of that data analysis serves as an information source itself, creating a stream of falsified data.

The most frequently used example to illustrate this is the case of Google Translate and Wikipedia. As they draw on many pairs of parallel texts from different languages, e.g. the same Wikipedia entry in two languages. However, for some less common languages the Wikipedia article itself might have been written using Google Translate. In that case, any error in Google Translate 'infects' also Wikipedia and is consequently fed back to Google Translate, reinforcing the error and thus, creating the 'echo'.

These limitations are important to consider when setting up the procedure of information collection and analysis through big data. However, there are two more concerns which have previously been mentioned, that become especially apparent during the application phase of big data standards within the city, namely cybersecurity and privacy.

Cybersecurity

Building a smart environment is reliant on the infrastructure provided through a network of data analytics across the entire area (Li et al., 2015). In other words, a city's services and spaces become dependent on software and funcion (Kitchin, 2013). While this might help cities in solving urban problems, it opens them in new ways, potentially creating vulnerability in other issues. Everyday, these software systems are becoming more complex and interconnected, which increases the challenge of producing stable, robust and secure devices and infrastructures (Kitchin, 2013). On top of these systems becoming more complicated, they themselves are progressively becoming open to malicious attacks. Also, the multiple devices, such as smartphones and tablets, that make up the internet of things are equally vulnerable to direct infiltration in form of information theft or damage to the user.

Despite these concerns, smart environments and networks of data analytics have been quite robust and so far have not become victims of such risks. Additionally, big data analytics bear opportunities for enhancing cyber security measures in an interconnected environment. These opportunities, however, have to take into account issues regarding individual's privacy.

Privacy

As more and more personal data are collected, the use of common practice analytics lead to growing privacy concerns. Protecting privacy becomes more difficult as the collected information is reproduced and distributed to a wide range of parties (Tene & Polonetsky, 2013).

Especially, the fleeting nature of personal data makes it difficult to recover information once it is exposed. The European Commission proposed a "right to be forgotten," which would allow individuals to demand from organizations to wipe their data slate clean (Tene & Polonetsky, 2013). On the other hand, this would increase the regulation of the internet which is why the proposal has been fiercely opposed by online platforms and free speech advocates.

Another matter is the issue of predictive analysis. While it is an extremely useful tool for law enforcement, national security, credit screening, insurance, and employment and would help enhance cyber security measures, it raises ethical dilemmas, similar to the small-number phenomena concern depicted above (Tene & Polonetsky, 2013).

Moreover, the benefits of big data tend to lean toward the government and big organisations and come at the cost of the individual. The importance lies in the fact that data extractors, governments, organisations, researchers, share the collected information and how they plan to analyse it, with the individuals that are the producer of that data. By ensuring a transparent process of big data analytics the privacy of the individual can be protected, while still exploiting the benefits of big data fully.

What these drawbacks and examples prove, is that while the benefits of big data analytics seem endless, the limitations are crucial in determining the area and scope of the potential application. Moreover, information retrieved through current big data standards alone cannot solve urban issues. As Glaeser et al. (2015) put it, big data will be able to support urban research far better when combined with something called "exogenous shock". This term basically describes the causality needed, together with the correlation, to be able to end up with verifiable results.

Rabari & Storper (2013) conclude that, the "digital skin" helps manage places better, but that this does not mean it will be able to "even the landscape of richer and poorer places.". The authors even proclaim that, more affluent places will have a greater ability to benefit from expensive digital technologies, widening the gap between rich and poor even further (Rabari & Storper, 2013).

Due to the immaturity of big data standards as a method in research, this study is mainly explorative. The identified potential opportunities are carefully reviewed against the crucial limitations and correlations are thoroughly examined before jumping to immediate conclusions. Albeit these above discussed limitations, it is believed that the possibilities of big data standards to enhance the management of urban dynamics outweighs the drawbacks and therefore, cannot be overlooked as a suitable means for information collection.

In order to fully grasp the potential of big data analytics with regard to unravelling city dynamics, the following chapter introduces various researches that have retrieved information through current big data standards, using various big data sources, to analyse urban areas with regard to socioeconomic characteristics and social networks of a neighbourhood.

3.2. Exploring Possibilities of Big Data Standards

This review of research on big data analytics with regard to unravelling city dynamics, shall illustrate the possibilities and advantages these standards pose compared to more traditional approaches and depict an overview of state of the art big data models.

Research by Dudas et al. from 2017 identified a correlation between socio-economic factors of neighbourhoods in New York and the distribution of Airbnb rentals within that area. According to the authors, Airbnbs accumulate in areas and neighbourhoods with certain features. By identifying which features and characteristics correspond with the aggregation of Airbnb rentals, interesting insights into the dynamics of Airbnb hosts and location decisions can be revealed. Moreover, the results of this study can also be used in reverse. By identifying areas and neighbourhoods with a large number of Airbnb rentals, conclusions can be drawn with regard to the neighbourhood's socio-economic profile. Ultimately, this opens the possibility of determining whether a neighbourhood, due to its socio-economic profile, is prone to be subject of neighbourhood change, such as gentrification.
A year earlier, Quattrone et al. (2016) focused on identifying the true beneficiaries of the advantages the sharing-economy of Airbnb claims to distribute. Again, The socio-economic characteristics of neighbourhoods were examined. This time, the analysis took place before and after the aggregation of Airbnb rentals and was compared to neighbourhoods experiencing no settling of Airbnb. Their findings show, that areas with a large number of Airbnb rentals were able to enhance their socio-economic condition over time. Conclusively, the amount of Airbnb rentals in an area can be seen as an indicator toward the likelihood of socio-economic neighbourhood change.
On the other hand Gurran and Phibbs (2017) researched the negative impacts Airbnb rentals can have on neighborhood characteristics. First, they examined how urban planners can react to the aggregation of Airbnb rentals within certain neighbourhoods. They established the impact of Airbnb rentals on a neighbourhood concerning issues such as noise, congestion, competition for parking, reduction in permanent rental housing supply and the increase in rental prices. Followed by actions that can be taken by policy makers and urban planners, to ensure neighbourhoods with a large amount of Airbnb rentals are able to cope with such arising issues. Through this research process it is again possible to identify neighbourhood characteristics, which are significant when assessing neighbourhoods and areas in preparation for urban (re)development projects.
Not only Airbnb can be used as an indicator of neighborhood characteristics. In 2012, Cranshaw et al. analysed the structure and composition of a city based on social media, in this case check-ins shared on Twitter, its residents generate. Identifying how areas are used and by whom, how they are characterized and what shapes them was the basis of their ambition to cluster this information and form meaningful 'livelihoods', which give insight into the socio-economic characteristics and demographics of a neighbourhood.
Also, the research by Boy and Uitermark (2015) looked at how a city's character is depicted in social media. By analysing user posts on the platform Instagram they identified how the city is presented

and perceived by residents and visitors. This social media platform is used to navigate the urban landscape and communicate belonging. Moreover, this analysis also revealed socio-spatial inequalities within the city.

Soliman et al. (2017) used a different angle when examining neighborhood characteristics through social media. The authors analysed Twitter users' mobility patterns to determine key locations of people within the urban landscape. Their approach is able to give a higher resolution on urban land use than contemporary collective tweet analysis, as individual activity streams are followed, rather than analysing groups of tweets at once. By retrieving 39 million geolocated tweets of specified users, travel surveys and parcel-level land use maps in the city of Chicago, they came to valuable conclusions with regard to determining which part of the city people prefer to use.

Additionally, to the character of places and how they are used, researchers have also focused on the user of the urban landscape. For example, Moseley et al. (2015) discovered the versatility of analysing linguistic patterns of tweets. According to the authors, demographic attributes of individuals can be determined through analysing tweet content. The information on gender, age or education level of a specific group of individuals is useful for purposes such as smart ad placement, security or identifying predator behaviour, but also for establishing neighbourhood profiles.

Likewise, the recent publication by Abitol et al. (2018) examined how key linguistic variables measured in individual Twitter streams depend on factors like socioeconomic status, location, time, and the social network of individuals. By focusing on determining the socioeconomic status of individuals from the way they tweet, the authors arrived at four major conclusions: socioeconomic indicators and linguistic variables are significantly correlated, the geographical location within the study area had influence on the results, standard language was more likely to be used during the day and non-standard language during the night, and the social network and status matters in terms of linguistic similarity between peers.

Not only analysing the current situation of neighborhoods, the usage and users is vital information. Also, how a city's characteristics are influenced by certain developments and how conclusions can be drawn from these with regard to housing stock and income levels, can be provided through big data. This is demonstrated by the study of Gant (2016), who points toward holiday rentals such as Airbnb, as the new gentrification battlefront. According to the author, the accumulation of Airbnb rentals and the attraction of tourists within certain areas leads to the displacement of residents. The results of this study therefore imply, whether a neighbourhood is exposed to displacement, hence likely to experience neighbourhood change.

The research by Yrigoy (2016) also describes the impact of Airbnb on the urban area. It is built on three assumption. Firstly, that Airbnb enriches the low-income population. Secondly, that Airbnb invigorates the housing market. Thirdly, that Airbnb boosts tourism in an area. No matter, whether the assumptions are correct or are disproven during the course of the research, this study still unveils multiple correlations between Airbnb rentals and socio-economic characteristics of an area. The findings eventually lead to the identification of the growth of income, the movement of the housing market and the popularity of an area.

Generally, determining the attractiveness of areas is a common aim of big data research. Bendler et al. (2015) traced the social media of individuals in the city to spot popular areas. The purpose of their study was to find better ways of city planning, venue recommendation and investment, through determining the attractiveness of places. Analysing the spatial and contextual characteristics of Twitter activity was core to their findings.

An unconventional way of determining how people experience the city was initiated by Paldino et al. (n.d.). The aim was to use information from geo-tagged photographs taken from Flickr to compare the spatial behaviour of residents and tourist within cities on global and local scales. Identifying tastes of individuals in the city, what attracts them to live in a particular city or spend vacation time there is important information for policy makers and urban planners.

While the above two studies focused on attractiveness for residents Sobolevsky et al. (2015) aimed at unravelling the attractiveness of places for tourists. For this reason they made use of the digital

information produced by human activity on the web to examine people's behaviour. In detail, the researchers were interested in determining which places in a city were most attractive to foreign visitors. By cross referencing bankcard statements with information from geo-tagged photographs from Flickr and Twitter they were able to arrive at insightful findings with regard to a place' attractiveness.

As opposed to determining negative impacts, Sans and Dominguez (2016) examined the advantages of Airbnb in the city of Barcelona. Airbnb is believed to revitalize and diversify neighbourhoods and provide economic support for families. Consequently, through identifying the number of Airbnb rentals, a neighbourhood profile can be established, expressing the state of liveliness and diversity and the economic status of families.

Very insightful on the interconnectedness of people and the area is also the study by Hristova et al. (2014). It is focused on identifying the social diversity of places through the Twitter network and Foursquare check-ins of residents and visitors, unveiling the social system of connections and tracking the mobility pattern. The study's aim was to determine places that bring together strangers as opposed to places that bring together friends, as this information discloses the level of social cohesion. Neighbourhoods with low levels of social cohesion tend to experience more impactful changes in neighbourhood dynamics.

The review of these researches shows through practical examples, the numerous possibilities of big data standards and the different big data sources to retrieve information valuable to managing urban dynamics. In order to ensure the validity of the retrieved information, big data results are often cross referenced with traditionally extracted data, such as census or survey data (Moseley et al.,2015; Dudas et al.,2017; Hristova et al., 2014). From this literature review also the two main big data sources used for research concerning urban social dynamics, can be identified. Most commonly applied are data retrieved from Airbnb and the social media platform Twitter. This supports the decision to focus on these two data sources further.

Below, the characteristics, opportunities and limitations of both, Airbnb and Twitter, will be introduced.

3.2.1. Airbnb Data

The Airbnb data mentioned in the studies above refers to the online marketplace and hospitality service offering overnight accommodation in 65.000 cities around the world, only indirectly. While the utilized information orginitates on this platform offering short-term rentals and providing booking services, the geolocated data is actually taken from the Inside Airbnb website (http://inside.aribnb.com/). Inside Airbnb is an independent, non-commercial initiative, which makes Airbnb data of more than 30 cities around the world publicly available for research. It was started by Murray Cox, who had the first idea for this website in 2014. The organisation clearly states that they are not endorsed by Airbnb nor by any competitor (Inside Airbnb, n.d.).

The Inside Airbnb website basically prepares the uncategorized information from Airbnb into data streams concerning the location, host name and ID, room type, price, minimum nights, number of reviews, listings per host and availability, to be mined directly from their platform for the purpose of research. This data is also made presentable through easily readable graphs and maps, making it even more manageable. Not only the large amounts of categorized data on Airbnb in cities, or the way in which the data is presented, but also the completely free access are reasons why this data source is so highly valued.

Naturally it also faces certain limitations. For one there is the possibility to come across outdated or even fake listings, so-called 'bait and switch listings'. Also it is not possible to determine whether properties have long been available for short-term rent (Gurran & Phibbs, 2017). Moreover, the terms and conditions under which the data is made available need to be strictly complied with in order to ensure the hosts privacy.

3.2.2. Social Media Data

The sheer range of smartphones and other portable devices makes the mobile access to social networks independent of time and location. From just about anywhere, a user is able to share his/her thoughts, feelings, photographs and place. Not only can users check-into a place they can also rate their experience there. As illustrated in the above examples this can give insight into the way a city is used and by whom as well as what people think about it. Twitter seems to be the prevailing candidate for extracting such information on the city.

Twitter is a platform offering online news and social networking services. It was founded by Jack Dorsey, Noah Glass, Biz Stone and Evan Williams in 2006. The platform allows users to post notifications and interact through messages called "tweets". Not only is Twitter one of the main sources from which individual's thoughts, feelings, photographs and location can be collected. Twitter can also be linked to other social media platforms such as Foursquare, a venue check-in and rating platform, and FLickr, an image and video hosting service. This means, that by mining one data source, the door is opened to also access information from others. There are many more social media platforms that produce potentially useful data outputs, such as Snapchat, an instant-messaging service, Instagram, an online photo and video sharing service, and of course Facebook, an online media and social networking service company, however due to strict regulations and inexperience with collecting data from these sources for research, they are not reviewed further.

Overall, these data sources bear certain limitations. First of all, data from social media platforms is generally unstructured. This means the data cannot immediately be categorized or analyzed nor can algorithms be applied. There is a large amount of work involved in pre-processing the information and making it viable.

Additionally, social media is often classified as inaccessible, meaning mainly a certain social group makes use of it, mostly individuals of middle or high income households between the ages of 12 and 50. Furthermore, the exact age group that uses a certain social media platform varies distinctly. While Snapchat and Instagram for example are frequented by the 12 to 25 year olds, Facebook has more users aged 25 and older, and Twitter is used across all age groups. However, older or poorer individuals might be excluded from using this technology entirely.

Moreover, social media is very dependant on its users. Most of the times, they are aware how the information they share on these platforms impacts others. Therefore, recommendations can be manipulated and reflect extremes (Bendler et al., 2015).

Awareness for these limitations, makes it possible to work with these data sources even better. For example, by taking a margin for outdated Airbnb listings into account or by considering that the collected information depicts the opinions of a certain group, meaning to represent the population entirely other means for collection need to be employed additionally.

Nonetheless it is believed, that both data sources are able to provide a useful basis for monitoring and managing the ongoings of a city or neighbourhood, as they open possibilities of retrieving information with higher resolution and in real-time.

From the knowledge gained through the above literature review on the possibilities that big data standards pose with regard to managing urban dynamics, such as introduced in chapter 2, a research procedure can be set up in order to explore these possibilities further. Through an investigative manner, the methods through which the introduced big data models can function as means for information collection, and thus support decision making of urban development and building construction projects, are critically reviewed.

The subsequent chapter introduces the research method, techniques and data proceedings.

4. Method

4.1. Means and Technique

As research concerning big data is quite young and not a lot of knowledge has been established around it yet, this study is primarily conducted by means of explorative research. It is aimed at broadening and deepening the knowledge base on big data standards, the limitations and opportunities, through exploring so far unasked questions and their possible answers. This is achieved through an extensive literature collection and review on two tropics, a case study analysis and the expert opinion of multiple housing associations.

In the previous chapters a literature review has taken place already. Firstly, concerning the questions of urban dynamics and the origin of inequality in cities in general and within the context of Amsterdam. From this literature analysis it was possible to derive the most essential indicators of urban dynamics with regard to gentrification in the Netherlands. This list of indicators represents the information currently considered during urban development and building construction projects and is guiding in formulating where big data analytics can be applied to better manage and enhance the outcomes of such projects in the future.

The second topic for which a literature study is conducted, concerns the examination of researches on utilizing big data analytics as a new method for information collection. Through this critical exploration it is possible to formulate assumptions regarding the opportunities current big data standards bear for the information collection on questions concerning urban dynamics. Moreover, adapted from the deployed methods and findings of these studies a blueprint can be built, representing how current big data standards can form a framework able to support the planning and decision making process of development questions concerning city dynamics, through information collection.

The case study allows for an illustration of this blueprint, making it possible to draw conclusions from this example with regard to the frameworks practical limitations and opportunities and its overall performance.

Finally, the outcome from the interviews with housing associations having ownership in the case study area, enable the framework to be examined from an organisational standpoint. Constructive implications and recommendations, based on stakeholder judgement, can then be added to the results.

In the following, the deployed methods and findings of the reviewed literature on the possibilities and limitations of big data standards are introduced.

4.2. Big Data Analytics: Advantages and Assumptions

In general, it has been established, that big data analytics has the advantage of making a considerable amount of information available without the need for information extraction through more traditional methods, that it enables real-time access to information and allows examination and analysis on a higher resolution. Big data analytics describe an unconventional new way, of information collection from social media sites and online platforms. The advantage of these sources is also that information can be linked to a place, so-called geo-locational data. This makes it possible to derive information embedded in a local context, taking complex city dynamics into account.

An extensive literature review has been conducted to authenticate the assumptions on big data advantages. The possibilities of big data analytics derived from this review have previously been introduced and the two main sources for information collection concerning questions of urban dynamics have been identified. In total, fifteen studies, six studies applying Airbnb data and nine studies applying social media data, have been reviewed, based on the researchers' expectations, the research method, limitations and added value for the purpose of this research. Below, the research methods and limitations will be elaborated on.

The methodology which Dudas et al. (2017) deployed in order to identify the correlation between socio-economic factors of neighbourhoods in New York and the distribution of Airbnb rentals, is a four step approach that combines both a big data method with traditional means.

First, the information on the amount of listings in an area is retrieved from Inside Airbnb. Next, the socio-economic factors of a neighbourhood are gathered through census data. Then, the attractiveness of the study area is determined from OpenStreetMap/geofabrik.de, a project initiated with the aim of providing free map data worldwide. The project focuses on mapping infrastructure, rivers and also points of interest. Finally, all this data is combined with the Pearson Cross-Correlation matrix, a statistical software. Apart from the general limitations the data source Inside Airbnb holds, such as the potential for reporting outdated listings, having no ability to determine for how long listings have been available for short-term renting, nor whether the reported listings are conventional Bed & Breakfasts now using the Airbnb platform for marketing and booking, also the traditional data source census data bears drawbacks, such as delivering possibly outdated information and requiring a quite extensive collection and analysis beforehand. While the data from OpenStreetMap/geofabrik.de might not be available in other study areas than New York.

In 2016, Quattrone et al. focused on identifying the true beneficiaries of the advantages the sharing-economy of Airbnb claims to distribute. Again, the starting point was the mapping of Airbnb listings per area from data retrieved from Inside Airbnb. This is followed by establishing the demand for Airbnb per area, using the amount of reviews per area compiled on Inside Airbnb, as a proxy for the demand. Next, socio-economic conditions of the study area are examined through census data giving insight into population density, the number of young people, the number of educated people, proxies concerning the pleasantry of areas and housing information on the number of flats and houses present, the number of properties sold, the number of dwellings owned/rented and the median house price. These socio-economic conditions are verified through a number of different tools. For one, the Gini-Simpson-Index, an ethnicity index indicating, whether two individuals chosen at random from an area are of different ethnicity. The Cultural Index, also bohemian index, representing the fraction of people employed in arts, entertainment and recreation. The Melting-Pot Index or foreigner index, showing the number of people born outside the country divided by the total number of people in that area. And the Index of Multiple Deprivation, which is comparable to a small level census, disclosing information on income, employment, health, education, the barrier to housing services, crime and the living environment. Finally, the attractiveness of an area is determined through the transport accessibility derived again from census data, through check-ins into venues within that area on the Foursquare plattform and through the Ordnance Survey, which captures areas that are covered with attractions, normalized by the size of the area in km2.

By correlating the information on the number of Airbnb rentals, with socio-economic indicators and the attractiveness of a place, the authors were able to determine, that the amount of Airbnb rentals in an area functions as an indicator toward the likelihood of socio-economic neighbourhood change.

Airbnb and census data behold the limitations mentioned previously.

Gurran and Phibbs' (2017) approach on how urban planners can react to the aggregation of Airbnb rentals within certain neighbourhoods, starts with establishing neighbourhood characteristics, such as noise, congestion or parking issues, through examining how tourism and residential activities are regulated, by reviewing state and local controls and regulations of certain areas. As in the studies above, the point-in-time information on the number of rentals in the area is established. As is the advertised availability, the number of Airbnb listings per host, nightly prices and reviews per month and composition of the listings e.g. whole houses, apartments, rooms or shared rooms through Inside Airbnb data. The data reveals information on the supply of rental housing and rental prices. Next, this information is correlated by reviewing rental vacancy rates from permanent rental advertisement data. Whether or not Airbnb rentals create an added income for their hosts is established by comparing census data in median monthly mortgages and state government data on median monthly rents with the average monthly incomes from each Airbnb accommodation type retrieved from Inside Airbnb.

Gathering all this information, results in the creation of a socio-economic neighbourhood profile. As this example makes use of the same data sources as the previous researches, it also faces the same limitations.

In 2012, Cranshaw et al. analysed the structure and composition of a city not through Airbnb, but based on social media data. For this purpose the researchers set up *Livehood* clusters, based on the characteristics, usage and users of an area. The clusters are established through aligning the check-ins from the Twitter public timelines, with the venue information retrieved through the Foursquare API, thus creating a cluster of information on users and the usage and characteristics of an area. Eventually, the researchers, validated their results through interviews.

The same limitations as above apply for data from Twitter.Generally, aggregating behaviour of many tends to hide behaviour of minorities. Moreover, a limited sample was used. All in all, the authors proclaim, that how an area really feels and what its social added value is can only be determined through experience. With regard to data from the Foursquare API, check-ins are usually generated by a demographic of young professionals between 25-35, who are owners of smartphones and urban residents, which can influence the results.

Repeatedly, Boy and Uitermark (2015) follow the approach of using social media data to determine, this time not the characteristics, but how a city is used. They utilize posts on Instagram, a platform for sharing moments in form of photographs, to discover how the city is restructured but also refractured through social media. First, it is determined in what way the city is used by analyzing Instagram photographs based on the content and representation of content. Secondly, the networks of users are revealed through the likes and comments on photographs. User communities, the segmentation of Instagram activity, are determined by looking at the most central account, their backgrounds, what images they circulate as well as the proportion of posts containing a tagged location, so-called spatial embeddedness. Finally, the level of inequality is reflected in the distribution of likes and comments (stratification). All this data put together indicates, the perception of the city through residents and visitors, the feeling of belonging as well as socio-spatial inequalities.

Limitations for data from Instagram include, the large amount of work required in pre-processing the unstructured data and the fact, that only certain population groups have access to this technology, similar to Twitter and Foursquare. In this particular case only public accounts with geo-tagged photos are used, limiting the scope even further.

Digging deeper into how certain individuals use the urban landscape, Soliman et al. (2017) analysed Twitter users' mobility patterns to determine key locations of individuals. For this purpose they developed a four scenario approach. The *Random Walker Scenario*, where geolocated tweets of a user group seem randomly located over the city scape when viewed over time, indicating tourist activity. The *Preferential Return Scenario*, where geolocated tweets of users cluster around key locations, e.g. home or work, when viewed over a longer time. The *Semantic Coherence Scenario*, in which the basic assumption is, that if the *Preferential Return Scenario* is the case, the key locations can also disclose the functions. Therefore, it can be assumed, that the top key location for users' tweets is home. The *Temporal Coherence Scenario*, which displays the possibility to assert each key location with the time of day based on the hours when the majority of tweets were posted. Assumably then, the timing of tweets is dependent on the land use type. Classification algorithms could be applied to infer land use types of key locations based on the timing of tweets at each location.

Apart from the previously introduced Twitter limitations, which include that the data is generally unstructured and the retrieved information is very dependent on the users, e.g. older or poorer individuals might be excluded from using this technology, the main limitation for this study is the challenge of a consistent data generation process.

A method used to determine not only characteristics and key locations, but also the user of the city is deployed by Moseley et al. (2015). The authors discovered the versatility of analysing linguistic patterns of tweets, by relying on previously selected abbreviation features by Gouws et al., to determine gender, age and education level.

Again the limitations for Twitter apply. Additionally, prior knowledge as to which abbreviations are the

essential characteristics of the to be determined demographics is necessary, which basically is a complete study on its own.

Also the recent publication by Abitol et al. (2018) examines the connection between linguistic patterns and socio-economic status of users. The authors analyse how key linguistic variables measured in individual Twitter streams depend on factors like socioeconomic status, location, time, and the social network of individuals. The approach begins by collecting a large Twitter corpus, including geotagged posts and proxy of social interactions of millions of users. This is followed by establishing a detailed socio-economic map, describing socio-economic indicators with a high spatial resolution in the study area. Then, three linguistic variables are extracted from individual Twitter timelines and studied as a function of the rate of standard negation, the rate of plural agreement and the size of vocabulary set. Finally, the information is processed via a detailed multidimensional correlation study.

The general limitation when using Twitter data apply in this case. It is known, that Twitter provides a biased sample. In this research a source of bias includes the "homogenization" performed by INSEE, which was deployed to ensure the privacy regulations are complied with and the self-devised proxies to approximate users' home location and social network are up-held. Moreover, pre-knowledge of the language in which one is analysing linguistic patterns is a requirement.

The study by Gant (2016) does not aim at characterizing the users of a city, but focuses on the impact of Airbnb on the users of the city. Once more, the amount of Airbnb listings in an area and the accommodation type of these listings from Inside Airbnb are the key information. The data is compared to the total housing stock as an indicators toward determining displacement within a neighbourhood. The data taken from Inside Airbnb experiences the same limitations as stated above.

Similarly, the research by Yrigoy (2016) analyses the impact of Airbnb on the urban area. The research method follows the three formulated assumptions. Firstly, it examines whether Airbnb influences the supply of rental housing, by reviewing the typology of units rented through Airbnb, as well as the hosts of these listings. The former is examined, as whole apartments for rent are likely to serve mainly as assets realizing the exchange value for property, instead of accommodating everyday life. The latter is examined, as the number of listings per host is indicative of private renters or corporations and rentals agencies. Secondly, the impact on real rental prices is determined through looking at the rental prices of Airbnbs in the area. Thirdly, it is analysed whether Airbnb creates an added income for individuals, the more listings from one host, the likelier it is that renting is run as a main business.The same limitations as described above apply to this research method.

As mentioned earlier, social media is a favoured method when researching attractiveness. Bendler et al. (2015) approached their research on city attractiveness from two angles. One, through determining the popularity of an area through the share of Twitter messages in that areas compared to all available Twitter message in the whole study area. And two, by determining the activity in an area through the sum of Twitter messages with additional content, e.g. a photo, video or linked place, divided by the total amount of Tweets in that area. By graphically overlapping the results of one and two the attractiveness of an area is determined.

As Twitter data is used for this purpose, the same limitations as above apply. Additionally, due to the fact that only tweets with geolocation are analysed, the scope of the research is very limited as only about 1% of released tweets through the Twitter API contain geolocational tags.

Paldino et al. (n.d.) follow an unconventional way of determining how people experience the city in terms of attractiveness. Based on the spatial distribution of the number of photos taken and posted on the sharing platform Flickr, the authors were not only able to determine the attractiveness of areas, but also whether it was attractive for residents or for tourists. The higher the picture density, the more attractive a place, while looking at the users profile discloses the amount of posts per user at a location over a given moment in time, indicating the residency status.

Generally, this approach is limited through excluding all photographs where no origin can be determined. Again, Flickr belongs to the type of social media frequented by a certain kind of social group, meaning only the attractiveness to some residents and visitors is determined in the end.

Likewise, the method deployed by Sobolevsky et al. (2015) foresees the assessment of the level of

attractiveness to tourists through the amount of geo-tagged photographs from Flickr. This information is correlated with the amount of geo-tagged messages from twitter, also indicating attractiveness to tourists. Finally, the validation of this method takes place through reviewing bank card transactions of foreign banks in certain places of the study area.

The limitations for Twitter and Flickr data correspond to the above mentioned limitations. With regard to the bank card transaction data, this was only retrievable as the researchers were cooperating with a main bank and had special agreements set in place, making this approach not transferable.

The positive effects of Airbnb are highlighted by Sans and Dominguez (2016), who examined the urban impacts of Airbnb on the city of Barcelona. First, the authors aimed at determining whether revitalization was taking place through Airbnb by overlapping Airbnb locations with hotel locations, to discover, whether Airbnbs are located in areas without or less hotels. By comparing the amount of rentals in an area with the migration rate taken from Census data and through examining listings' descriptions to determine the genuity of the offered experience, this indication is made viable. Through the examination of host origins, by comparing host languages to the population groups registered in the area from a country speaking that language, the diversification of the neighbourhood is established. Whether a growth of income through the Airbnb supply can be registered and the average income of that area generally, are determined through the characteristics of family households and the number of listings, as well as the characteristics of listings offered within the area. Data retrieved from Airbnb and census faces the drawbacks as described above.

Complex but insightful on the interconnectedness of people and place, is this study by Hristova et al. (2014), who arranged a method to identify the social diversity of places. The authors began by determining the social network of individuals through their direct Twitter network: if user A follows user B and user B follows user A back a social network is established. Secondly, the place network was determined through the check-in history of individuals on Foursquare: if user A goes to venue X and then to venue Y and user B does the same, then a place network is established. This is followed, by determining the type of venues within the placenetwork and their geographical locations from Foursquare. Next, areas of deprivation are pointed out through the Index of Multiple Deprivation. By correlating the information on venues, bridging places and bonding places are revealed. The former being places that bring together stranges (brokerage), the latter places that bring together friends (serendipity). Correlating the information on users and deprivation reveals the situation of urban social mix of an area (entropy and homogeneity), which ultimately is an indication for the level of social cohesion. The generally limitations of Twitter apply here.

Finally, from this extensive analysis of big data methodologies, and taking into account the limitations of each approach, it can be determined which of these methods are eligible to be pursued further in the context of this research. Below a table depicts each of the analysed studies, for Airbnb data as well as social media data, stating the purpose of the study, the research expectation and most importantly, the applicability of the big data method for this research (Table 3 and 4) (see Appendix A - Big Data Assumptions for more details).

Social Media Research	Expectations	Applicability
1. Hristova et al. (2014) Social diversity of urban localities through the social network and mobility patterns of their visitors	Identifying social cohesion (places that bring together strangers vs. places that bring together friends)	YES
2. Cranshaw et al. (2012) Studying the structure and composition of a city based on social media its residents generate	Identifying how areas are used and by whom, how they are characterized and what shapes them	YES
3. Moseley et al. (2015) exploring demographic attributes of social media authors for purposes such as smart ad placement, security, predator behaviour etc	Identifying gender, age educatinon level of a individuals in a neighbourhood	NO
4. Boy & Uitermark (2015) restructuring and refracturing og the city through social media, using social media to present, perceive and navugate the urban landscape and communicate belonging	Identifying how the city is perceived, used and represented and uncover socio-spatial inequalities	NO
5. Bendler et al. (2015) spotting popular areas in the city through tracing social media of its users with the purpose of better city planning, venue recommendation and investment	Identifying the attractiveness of an area	YES
6. Paldino et al. (n.d.) unconventional way of how people experience the city using information from geo-tagged photographs, comparing spatial behaviour of residents and tourist on global and local scales	Identifying tastes of indivduals in the city, what attracts them to live in a particular city or spend vacation time there	YES
7. Sobolevsky et al. (2015 Using digital information produced by human activity to examine human behaviour	Identifying city attractiveness for foreign visitors	YES
8. Soliman et al. (2017) analysing users's Twitter mobility patterns to determine key locations of individuals which give a higher resolution on urban land use than collective tweet analysis	Identifying key locations and how the urban landscape is used	YES
9. Abitol et al. (2018) how key linguistic variables measured in individual Twitter streams depend on factors like socioeconomic status, location, time, and the social network of individuals	Identifying socioeconomic charcteristics of tweeters through analysing linguistic variables in theri tweets	NO

Table 3 - Expectations and Applicability of Social Media Data

Airbnb Data Research	Expectations	Applicability
1. Dudas et al. (2017) How socioeconomic factors of neighbourhoods impact the distribution of Airbnb rentals	Indentifying **neighbourhood characteristics**	YES
2. Gurran , Phibbs (2017) Impact of Airbnb rentals on neighbourhoods concerning noise, congestion, competition for parking, reduction in permanent rental housing supply, increase in rental prices	Indentifying **neighbourhood characteristics**, **rental housing supply** and **increase in rental prices** and **growth of household income**	YES
3. Yrigoy (2016) Impact of Airbnb on the urban area; 3 assumptions: Airbnb enriches low-income population, Airbnb invigorates housing market, Airbnb boosts tourism	Indentifying **income/growth of income**, **impact on housing market**, **social structure** of the area (popularity)	YES
4. Quattrone et al. (2016) Who benefits from the sharing economy of Airbnb	Identifying **socioeconomic conditions** of areas benefitting from Airbnb	PARTLY
5. Gant (2016) Holiday rentals: The new gentrification battlefront	Indentifying areas of **resident displacement** through hioliday rentals attracting tourists	YES
6. Sans, Dominguez (2016) Unravelling Airbnb: urban perspectives from Barcelona; Urban impacts of Airbnb on revitalizing nieghbourhoods, increasing and diversifying neighbourhoods, providing economic support for families	Identifying **urban impacts** of Airbnb on neighbourhoods and individuals	YES

Table 4 - Expectations and Applicability of Airbnb Data

This evaluation also highlights what is critical about some of the approaches. As mentioned, a strong limitation of social media is the fact that it is inaccessible to some social groups. Apart from that, the connection made between two individuals following each other and the impact of a social network is also questionable. Is it truly a given that two individuals, followers in the virtual world, are also social companions in reality? Moreover, some corroborations seem far fetched, such as the correlation between linguistic patterns or tweet content and the socio-economic status of individuals. Not only is a lot of prior knowledge necessary to come to such conclusions, also it is important to take into account that social media promotes the use of slang and abbreviations and that content can easily manipulated by the user not necessarily reflecting their true age, gender or education level. Or the relationship between languages listed in host profiles and host origins. A Spanish speaking host, does not necessarily have Spanish origins and even through cross referencing this data with information on Spanish foreigners living in the area won't guarantee exact composition of diversity.

Nonetheless, lining up the expectations and determining the applicability of the above big data methods, makes it possible to derive general assumptions, with regard to the possibilities of Airbnb data and social media data.

Overall, data collected from Inside Airbnb is able to reflect information on the people, such as the growth of household income, the property, such as rental housing supply and increase in rental prices and the area, such as neighbourhood characteristics, the social structure, socioeconomic conditions, resident displacement and urban impacts of Airbnb on the vitality and diversity of neighbourhoods.

Social media data on the other hand, can give valuable insights into social cohesion - based o the assumption followers are truly friends -, how areas are used and by whom, key locations of the urban landscape and attractiveness to residents and tourists, which is all information concerning the area, however, it only reflects the opinions of a certain user group.

These assumptions are critically reviewed and translated into concrete big data indicators. In order to enhance the management of urban developments and building construction projects, information on

these indicators is collected through the introduced big data standards and employed in the planning and decision making phase.

Table 5 below depicts the big data indicators within the three categories people, property and area, while an icon indicates the data source.

	Big Data Indicator	
People	growth of houshold income	airbnb
Property	rental housing supply	airbnb
	increase in rental housing prices	airbnb
Area	neighbourhood charcateristics	airbnb
	vitality and diversity	airbnb
	attractiveness	twitter / flickr
	key locations	twitter
	usage of urban landscape	twitter / foursquare
	social cohesion	twitter / foursquare
	resident displacement	airbnb

Table 5 - Big Data Indicators

This table shows the information current big data standards can assumably provide. Especially, in the category area the value of big data analytics to extract information on a very detailed, local level is demonstrated.
From the extensive literature exploration it was also discovered that the above introduced indicators are not collected through traditional methods and thus, so far not taken into account in decision making processes of urban development and construction projects in the built environment. The following chapter will introduce how information on these indicators can be provided for the purpose of managing city dynamics.

4.3. Extracting Information on City Dynamics through Big Data Analytics

The city is a dynamic organism consisting of multiple layers. The spatial component, such as the environment, the physical component, such as the buildings and the social component, such as the population. All three are subject to the dynamics coming with neighbourhood change, e.g. through urban development and building construction projects or gentrification processes. This is also reflected in the parallels that can be drawn with Fenton's (2016) definition of gentrification. The city's layers correspond with the categories named in the definition.
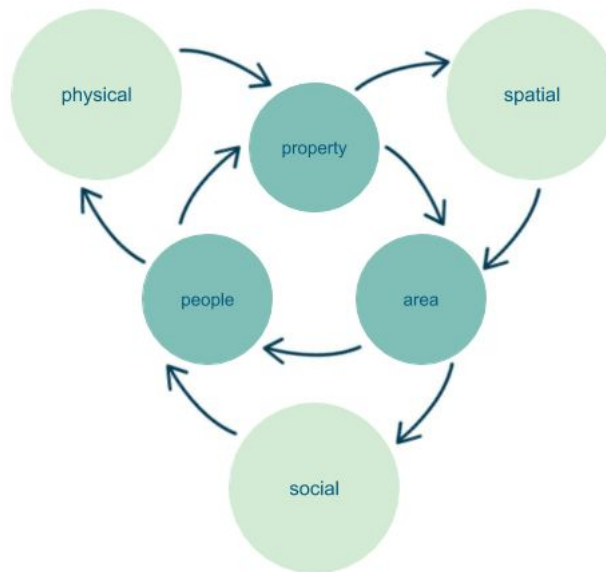
Figure 4 - Link City Layers and Gentrification Categories

The identified big data indicators have the potential of adding additional information to all three layers of the city, but especially the spatial component, as per Fenton's (2016) definition the area category. This information is considered essential to decision making processes, as it offers a detailed context and contributes to a holistic neighbourhood profile.

The below scheme (Figure 5) illustrates how the collection of information on the big data indicators is interconnected and should be integrated to provide information for a holistic neighborhood profile. From left to right the data sources, the big data indicators and the big data parameters, meaning the details that are considered to retrieve the actual information, are depicted. The connections between them show which indicators are derived from which sources and which big data parameters are needed to collect the information on the big data indicators.
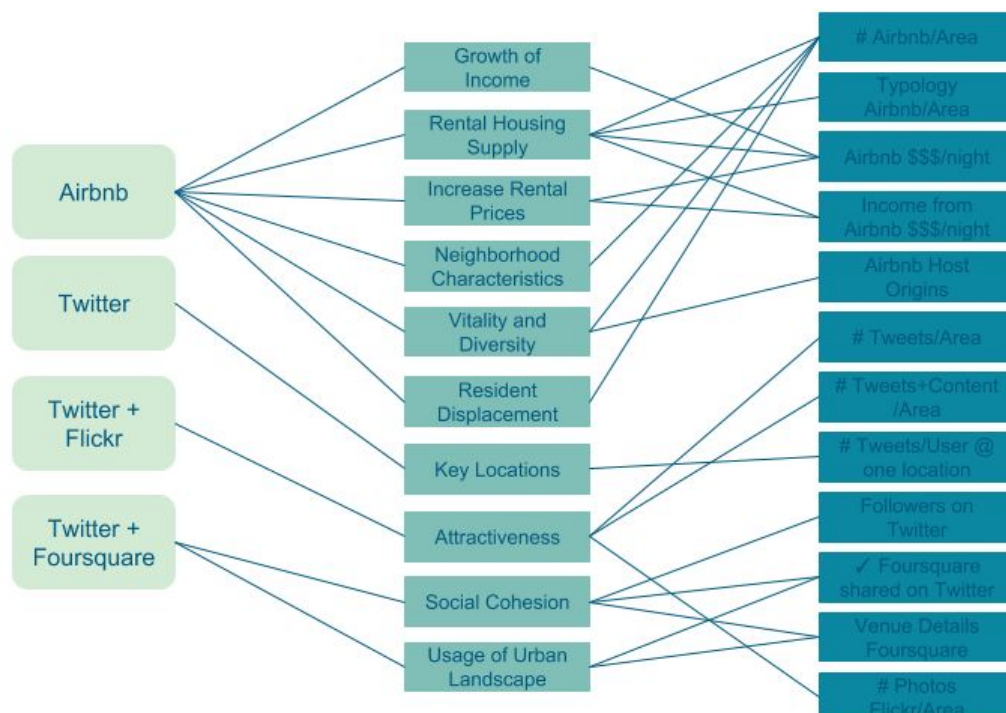
Figure 5 - Interconnectedness of Big Data Indicators

What the scheme expresses is not only, one big data parameter can give input to multiple big data indicators, but also, that the indicators require the input of multiple big data parameters.

Information on the big data indicators extracted from Inside Airbnb, such as *growth of income, rental housing supply, increase in rental prices, neighbourhood characteristics, vitality and diversity* and *resident displacement* are analysed through big data parameters, such as the number of Airbnb rentals per area *(#Airbnb/Area),* the typology of Airbnbs per area *(Typology Airbnb/Area),* the average nightly price of Airbnb rentals *(Airbnb $$$/night),* the average income generated from Airbnb *(Income from Airbnb $$$/night),* which is the average nightly price times the number of nights per year the rental is let, and the origins of the Airbnb hosts *(Airbnb Host Origins).* Twitter alone can give insight into *Key Locations* of urban areas through input from the big data parameter following the number of tweets an individual user has posted within the vicinity of certain locations *(#Tweets/User@OneLocation).* Twitter, and Twitter data in combination with Flickr, is able to produce data on the *Attractiveness* of an urban locality, with the input of the big data parameters representing the number of tweets found in an area *(#Tweets/Area),* the number of tweets containing an added content, such as photograph, video or geo-tag *(#Tweets+Content/Area)* or the number of photographs shared on Flickr within an area *(#PhotosFlickr/Area).* Twitter and Foursquare together reveal data on the indicators *Social Cohesion* and *Usage of the Urban Landscape* though big data parameters, such as who follows who on Twitter *(FollowersOnTwitter),* the check-ins on Foursquare which are shared on the Twitter public timeline *(✓ FoursquareSharedOnTwitter)* and the details on venues retrieved through the Foursquare API *(VenueDetailsFoursquare).*

Below, the detailed steps of information collection are depicted per big data indicator.

Growth of Household Income
Generally, the growth of household income is influenced by the Airbnb supply of an area and the average income of that area, the characteristics of family households in that area and the number of listings and characteristics of listings offered by the host in that area (Sans & Dominguez, 2016).

More concrete, through comparing census data on median monthly mortgages and state government data on median monthly rents with the average monthly incomes from each Airbnb accommodation type the added income is identified (Gurran & Phibbs, 2017).

It needs to be taken into account, that more listings per host do not directly point to a bigger added income, but more toward the Airbnb rentals being run as a business (Yrigoy, 2016).

Rental Housing Supply
In Gurran and Phibbs (2017) the supply of rental housing is determined through examining the amount of airbnb rentals in an area. This data is extracted from the independent website insideairbnb.com and includes detailed data on point-in-time information on the number of rentals in the area, the advertised availability, the number of Airbnb listing per host, nightly prices and reviews per month and the composition of the listings e.g. whole houses, apartments, rooms, shared rooms. data. Specifically the typology of listings rented and the owners or hosts of listings and the number of units they offer, is pivotal. As the amount of Airbnb rentals in an area grows, the amount of common rental housing declines. Apartments and houses are permanently taken off the regular housing market to flexibly accommodate visitors throughout the year. Especially, with whole apartments for rent it is likely that their main purpose is not everyday life but realizing the exchange value of the property. Additionally, it is more likely that one host with many listings is actually a corporation or a rental agency (Yrigoy, 2016).

Results from this data analysis are correlated with rental vacancy rates from permanent rental advertisements data, to ensure validity (Gurran & Phibbs, 2017).

Increase in Rental Prices
The data collected above for the purpose of determining the supply in rental housing is also indicative for the impact Airbnb has on rental prices. Here, particularly the nightly prices of Airbnb rentals are

significant. The higher the nightly fee, the higher in return the monthly rental price of common apartments and houses will be, in order to maintain profitability for the owner, who is letting property in the traditional manner to permanent renters (Gurran & Phibbs, 2017; Yrigoy, 2016).

Neighbourhood Characteristics

Identifying neighbourhood characteristics of an area starts by determining the amount of Airbnb rentals in an area and determining the impact they have on the urban landscape (Dudas et al., 2017; Gurran & Phibbs, 2017). This is done by then cross-referencing this information with data on socio-economic characteristics derived from census data (Dudas et al., 2017) or by examining how tourism and residential activities are regulated through reviewing neighbourhood characteristic such as noise, congestion or parking issues (Gurran & Phibbs, 2017).

Vitality and Diversity of Neighbourhoods

Analysing the process of revitalization of a neighbourhood is achieved through determining locations where Airbnb rentals are overlapping with traditional hotel locations. Airbnb locations in areas with few or even without any hotels are considered as 'area boosters', attracting a various set of individuals to the neighbourhood. Also, the listings description is examined to determine the genuity of the experience in that that area (Sans & Dominguez, 2016).

The diversity of individuals permanently located in a neighbourhood is examined through Airbnb host origins. By comparing the stated host language in the Airbnb profile to population groups speaking that language and that are registered in the area, the cultural mix of individuals can be identified (Sans & Dominguez, 2016).

Resident Displacement

Displacement in a neighbourhood is uncovered through examining the amount of Airbnb listings in an area compared to the total housing stock. Additionally, analyzing the type of listings offered in an area gives insight into the amount of housing permanently diverted from the regular housing market, resulting in resident displacement (Gant, 2016).

To corroborate results, the amount of rentals in an area is compared to the migration rate taken from census data (Sans & Dominguez, 2016).

Key Locations

Identifying key locations requires the close examination of individuals' Twitter activity. Users' geo-tagged tweets are examined closely and clusters of tweet activity from certain locations are identified. By correlating users' top tweet location with the time of day at which the tweets are posted, can give insight into the function of that location, consequently revealing how certain individuals gather around specific spaces within the city (Soliman et al., 2017).

Attractiveness

First, the popularity of an area is determined through the share of Twitter messages in that area compared to all available Twitter messages in the whole study area. Then, the activity in an area is determined through the sum of Twitter messages with additional content, such as photos, videos or linked places, divided by the total amount of Twitter messages in that area. Finally, the attractiveness of an area is determined by overlapping the mapped results of popularity and activity. Concluding, that high popularity and high activity make a place more attractive (Bendler et al., 2015).

In order to distinguish between attractiveness for residents and visitors, it is necessary to analyze the publisher of tweets either through their profile or through the amount of tweets per user at a location over a given amount of time (Paldino et al., n.d.).

Not only Twitter messages and photos can be a source for the level of attractiveness. Also on the platform Flickr users share their geo-tagged photos from which the same valuable insights can be drawn as from Twitter data, given that it can be distinguished who took the photograph, a resident or a

visitor (Sobolevsky et al., 2015).

Social Cohesion
Determining social cohesion in a neighbourhood through social media takes a four step approach. First, the social network of individuals is determined through their direct Twitter network. If A follows B and B follows A back there is a social connection. Second, the place network is determined through the check-in history of individuals on Foursquare. If A went to X and then to Y and B and C have the same movement pattern there is a place network. Thirdly, the type of venue and its geographical location are identified through Foursquare characteristics. Next, areas of deprivation are determined through the Index of Multiple Deprivation (IMD). Finally, by correlating the information on venues such as their place network, type and geographical location, bridging and bonding places are revealed, which function as indicators of brokerage and serendipity. Correlating the information on users and deprivation areas reveals the urban social mix to be found, which indicates entropy and homogeneity. Brokerage, serendipity, entropy and homogeneity together indicate the level of social cohesion of an area. There is a tendency, that the more diverse and deprived, the more prone an area is to gentrification (Hristova et al., 2014).

Usage of the Urban Landscape
Through aligning the user check-ins from the Twitter public timeline with the venue information on usage and characteristics collected from the Foursquare API, 'Livehood clusters' reflecting the characteristics, usage and users of an area are determined. This data does not only show who uses areas within the city, but also how certain spaces are used, e.g. for nightlife, daily shopping or culture. Overall, the results depict a feel of the neighbourhood in terms of what to do there and the people that can be found there (Cranshaw et al, 2012).

Knowing the workings of the information collection per big data indicator is the preparing step toward forming a blueprint from these big data standards. Also, this examination demonstrated, how big data standards are often correlated with information retrieved through traditional data sources to mitigate some of the limitations, such as excluding social groups from being represented or correlating data wrongfully. Therefore, also the blueprint will combine information from big data sources as well as from traditional sources. In other words, also the traditional indicators for the Netherlands introduced in chapter 2 are taken into account. That way, an integrated framework is prepared, to ensure, that a truly holistic neighborhood profile can be retrieved through this approach.

Before introducing the integration of the big data standards and the traditional methods, the following chapters contain the data plan and ethical considerations, as well as the performance of the developed framework, represented through an illustration in the case study area and how its managerial added value is planned to be substantiated through expert input.

4.4. Data Plan

Due to the delicacy of the information, it is important to study how the data will be handled during and at the end of this research study.

Despite its sensitivity, social media data is often classified as being openly accessible, as the information is made available through application programming interfaces (APIs) (Kelley et al., 2013). This means a social media data source, e.g. Twitter, sets up APIs, which specify how software components should interact with each other, in other words which parts of information are actually openly accessible. Twitter chooses which parts of the data not so share openly, based on their business interests, but also privacy reasons and ethical considerations. This results in numerous restrictions for researchers wanting to apply data from Twitter in their research. The major restrictions important for this research include, e.g. that the access to full tweets is limited. The APIs of Twitter are set at reliving about 1% of all tweets created, how these are selected is unclear. Additionally, the

possibility of reviewing tweets concerning a certain criteria, is limited through the accessibility of only very recent tweets. Moreover, there is a maximum of tweets that can be requested for a certain individual's twitter history (Kelley et al., 2013.).

Also on Foursquare Data the release of information is regulated through API. Moreover, there are strict guidelines regarding the usage and saving of data.

Usage of Foursquare data is only allowed if the data application is non exclusive, publicly available, without cost or subscription, does not cross the line of daily invocation and is non commercial. Apart from that, the Foursquare data on venues is under no circumstances to be saved for more than 30 days. The data on user check-ins no longer than 24 hours and data containing specific user information no longer than 3 hours. Furthermore, monitoring the course of check-ins of one individual user or saving this check-in data is only permitted when that user has been informed regarding the usage of this kind of data and his explicit allowance has been granted (Foursquare, 2017).

Data derived from Inside Airbnb, consists of truly openly accessible data from the Airbnb site. No 'private' information other than publicly available data, such as names, photographs, listings and review details, is being used. Moreover, location information for listings is anonymized by Airbnb. The location of a listing as seen on a map or in the data will be about 150 meters away from the actual address. Listings within one building are also anonymized individually and therefore seem spread around the actual address (Inside Airbnb, n.d.).

The above shows, that handling this sensitive data with care, is a topic within all the three data platforms. They have taken action and developed procedures for data handling and storage that either prepare the data in the right way, through anonymization, or limit the released data amount to a specific selection of information and at the same time make it impossible to store it privately.

4.4.1. Ethical Considerations

A big part of the data plan are ethical considerations that require attention when handling sensitive information. Here, information privacy is one of the biggest concerns. The ability to isolate or entirely withhold information about oneself is central to the idea of privacy. However, in the big data age it becomes difficult for the individual to monitor what information is being shared and with which parties. Additionally, the subjects and areas considered private differ among cultures and individuals (Le Roux, 2012). Inadequate or non-existent disclosure agreements need to be avoided at all cost. Otherwise, this will not only damage individuals, but also, the full potential of big data cannot be exploited. As seen above, the data platforms are as well concerned about their users' privacy and are taking adequate precautions.

A second concern is the fair usage of the information when it is released. Once the data arrives at the researcher, it is necessary to follow the FAIR guiding principles. These principles are leading in making sure data publishers act in accordance with the implementation standard of being Findable, Accessible, Interoperable and Reusable.

Apart from that, researchers are responsible for ensuring non-commercial data usage and for allowing the analysed data to be used for public analysis, discussion and community benefit.

Looking more specifically toward data taken from geolocational platforms, such as social media data, most users are aware of the information they are disclosing. This is also true for information disclosed on the Airbnb site. Individuals set up their profiles and agree to make it publicly accessible to others. Nevertheless, the process of data collection and data analysis must be transparent and on request accessible to all individuals to whom the data is connected.

4.5. Illustration of the Framework's Performance

The integrated framework is built on assumptions derived from extensive literature explorations. While through this literature exploration sufficient ground is established for investigating the added value of

big data analytics for managing and enhancing urban development and constructions projects in the built environment, this calls for an illustration of the blueprint's applicability. The aim of the framework is to add high resolution and real-time information on complex local networks at the neighbourhood level to the current information collection and ultimately deliver a holistic neighborhood profile. Together the big data indicators and traditional indicators will potentially be able to reveal data on urban dynamics which are essential for policy makers, urban planners, developers, strategists or anyone involved in the planning and decision making of urban development projects.

The illustration will take place based on a case study analysis. The selected area has recently been undergoing radical changes in the composition of social structures as well as building developments. Despite detailed planning and decade-long implementation, critiques are voiced regarding the development outcome. The framework will be used to authenticate, whether these critical observations are factual and can be confirmed through big data standards. The result of this illustration will be the formulation of proposals, stating where already during the planning and decision making process the framework could have been employed to mitigate the negative consequences that have arisen now.

The possibility, that the framework might fail to perform, meaning, that the observed critiques cannot be confirmed, could occur due to different reasons. For one, the criticism could be non-confirmable, as it is sincerely not true. Or, the retrospective manner of the illustration is not able to obtain conclusive results as the framework was designed to predominantly support planning processes. Or simply, the theoretically developed design of the blueprint and the promised approach through big data standards cannot hold what is expected, when translated from theory to practice.

Therefore, to determine the anticipated ability of the integrated framework to perform, the above described illustration takes place in chapter five.

4.6. The Framework from an Organisational Standpoint

Finally, should the framework's performance turn out successful, it will also be assessed through stakeholder judgement, meaning the standpoints of housing associations with ownership in the case study area. Their opinion is deemed relevant as by their ownership, they are seen as key stakeholders in the revedelopment process.

The first contact is established to confirm the indicators on people, property and area that the housing associations are interested in and are collecting for their planning purposes. Secondly, housing associations are asked to identify variables outside of their current collection but that would be relevant to them. And thirdly, it is asked for their opinion with regard to social media data as an empowering instrument for associations (see Appendix B - Survey Housing Corporations).

The companies are contacted by email or through their websites and asked to fill in a form with the requested information. This short survey has two purposes. For one, confirming from the organisational side the gap that has been determined in collecting indicators. And secondly, whether housing associations in the roles of planners see an added value in an information collection approach that uses social media as a data source. Do they in the future see the necessity of appointing a data manager or even establishing a data center within their organisation? Are they aware of the advantages and the power big data standards could add to their position?

For this procedure, the housing associations are asked to fill in a table with regard to the information variables they value and their ability to collect these variables. Additionally, the option to add additional variables they do already collect, is offered. Furthermore, two open ended questions are asked. The first with the aim to establish what information, apart from socioeconomic, housing associations might value. Ultimately, the desired response would include information variables on the field 'area'. The second question aims at the organisations standpoint toward social media as an information collection source. They are free to write their opinion on the topic of the added value of social media, appointing a data manager or setting up a data analytics department.

The outcomes from the survey will be added to the results of the framework's performance, to be able to formulate comprehensive and practical recommendations. Following this first contact the parties are invited to participate in an idea exchange and discussion sessions, during which practical matters and requirements for the implementation of such an integrated framework in the planning and decision making process in the future, will be discussed.

5. Case Study

5.1. Introduction to the Case Study

Along with the other research techniques, the illustration of the framework's performance within a case study analysis supports the research in discovering the added value of big data standards for planning and decision making processes.

For this purpose, Amsterdam has been selected as the study area. Not only the complex gentrification policies in the Netherlands and the unique history of gentrification in Amsterdam specifically, have been reasons for this choice.

Another motive is the fact, that Amsterdam is more mixed than it used to be and is repeatedly termed a 'super-diverse' city (Zwiers et al., 2016; Savini et al., 2015). While this creates an outstandingly various environment greatly suitable for this research, it also brings along more poverty concentrations. Low-income neighbourhoods in Amsterdam are noticed to increasingly concentrate outside the city centre (Zwiers et al., 2016). This creates spatially and socially disadvantaged peripheries (Kühn, 2014), while the inner-city neighbourhoods are maintaining their high status, supported by gentrification policies (Zwiers et al., 2016; Savini et al., 2015). This makes Amsterdam a preferred area to study this complex urban issue.

Moreover, the many regeneration projects within and around the city, as part of the long-term Amsterdam structural vision 2040[1], make it possible for further research to illustrate the framework performance directly to ongoing redevelopment projects and determine where and how big data standards could possibly enhance the management of such ventures.

For this illustration, one such regeneration project has been selected on the basis of certain criteria. Amsterdam Nieuw West and its neighbourhoods Geuzenveld-Slotermeer, Osdorp, Slotervaart and Overtoomse Veld depict a large-scale urban renewal project of a post-war residential area. The area has been known for the lack of social cohesion and sense of place, in other words it was a former problem 'concentration area'. During the redevelopment, the close cooperation of all involved stakeholders, including the residents was key to the project initiators. Through state-led gentrification, the municipality of Amsterdam drove the redevelopment, with the aim of improving the neighbourhood's social, economic and spatial situation (EUKN, 2010; Nio et al., 2016). Since the project has recently been completed, it offers excellent grounds for illustrating the framework's performance by examining whether the municipality's goals were achieved and even if it could have offered vital guidance at the start of the project.

5.1.1. Case Study Description

The district Nieuw West lies in the west of Amsterdam bordering Bos and Lommer in the east, the Westpoort in the north, the Schiphol Airport in the south-west and the Amsterdam forest in the south (Figure 2). The first plans for a residential development in the area were formulated as part of the General Expansion Plan of Amsterdam after the Second World War. Most of the development plan was focused at expanding the inner city of Amsterdam westward, resulting in the Nieuw West or Western Garden Cities (*Westelijke Tuinsteden*) development. The original design was based on the 'Garden City' philosophy from the late 19th century. The development concept by urban planner Cornelius van Eesteren foresaw a balanced mix of light, air and space (ISHF, n.d.; van Esteren

---

[1] The Amsterdam structural vision 2040 forsees the implementation of six spatial tasks in order to gather economic strength, become sustainable and be a driving force within the metropolitan context. For more detail on the projected interventions as part of the long-term plan visit
https://www.amsterdam.nl/bestuur-organisatie/organisatie/ruimte-economie/ruimte-duurzaamheid/making-amsterdam/structural-vision/

Museum, n.d.) and was presented as a progressive and modern living environment 'for the new Man' (MO, 2013).
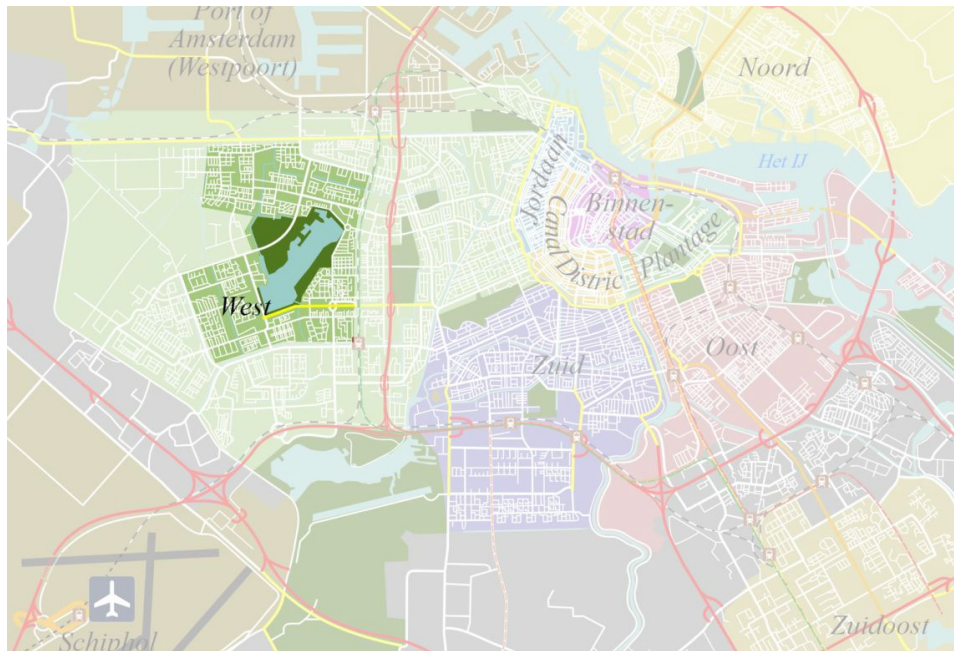
Figure 6 - Location Nieuw West within Amsterdam Metropolitan Region

However, this design concept was overthrown in the last planning phase, as the urgent need for more housing was voiced (ArchEX, n.d.). Thus, the originally planned, rather low density, single family homes and the garden city concept were largely replaced by the so-called concept of *Strokenbouw*, which are large free-standing blocks of flats (MO, 2013) (figure 6).

As mentioned, the Nieuw West district contains multiple smaller neighbourhoods with quite distinguished characteristics. Below a brief profile of the neighbourhoods in Nieuw West.

Figure 7 - Strokenbouw in Amsterdam Nieuw West

Geuzenveld-Slotermeer

Geuzenveld and Slotermeer (Figure 8) were farmland until as late as the 1950s. Then, the area was developed for residential use as part of the Amsterdam expansion plan. However, the general construction of houses was poor due to the severity of the post-WWII period (iamsterdam, n.d., A).

Before the consolidation of Geuzenveld and Slotermeer and the allocation of both from the Amsterdam district to the Nieuw West district in 2010, they were separate neighbourhoods, with different housing concepts and development focus, designed by different architects and realized several years apart. The uniqueness of each is tried to represent in the following.

The garden city of Geuzenveld was realized during the mid-fifties. It actually comprises six smaller neighbourhoods situated around the main square Lambertus Zijlplein. Each of these small neighbourhoods was designed by a different architect. However, the aligned goal was to create functional living spaces and ensure simple and thought through buildings. As the dwellings originated from very different architectural ideas, Geuzenveld turned out as a very diverse neighbourhood (vanesterenmuseum, n.d., A).

Figure 8 - neighbourhoods Geuzenveld (left) and Slotermeer (right)
(source: adapted from http://vaneesterenmuseum.nl/en/garden-cities/geuzenveld/ and
http://vaneesterenmuseum.nl/en/garden-cities/slotermeer/)

Slotermeer is the oldest of the Garden Cities. It was inaugurated by Queen Juliana on the 7th of October 1952. Shortly afterwards the first residents arrived either from the overcrowded city centre or as migrant workers from the countryside. The neighbourhood offered plenty of space, greenery and water, which created a pleasant living environment. At the heart of Slotermeer, the lively Plein '40-'45 is surrounded by many shops and local businesses, the 'Garden City Hall' (the district office since 1990), hotels and a small harbour. The square is also home to the Freedom Carillon which, like the name of the square, commemorates the Second World War (vanesterenmuseum, n.d., B).

Osdorp
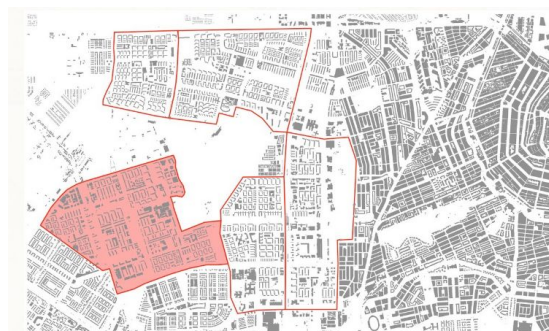
Figure 9 - Osdorp
(source: adapted from http://vaneesterenmuseum.nl/en/garden-cities/osdorp/)

In contrast to Slotermeer, Osdorp is the newest of the garden cities. Nonetheless, in recent years the 'East Village' has become a central point for people in the surrounding neighbourhoods. Mostly

constructed in the 1960s, Osdorp has the most striking skyline, especially when viewed from the Sloterplas lake. The high-rise constructions featured in Osdorp function as aesthetic accents among the neighbouring low-rise developments, as well as meet the steady demand for housing. This move towards high rise buildings brought in new designs, for example, the apartment buildings on the Notweg, which, at the time, were considered a novelty, as they were equipped with elevators, accessible ramp ways and built in racks to hide the laundry on the balcony (vanesterenmuseum, n.d.; C). At the same time, the focus was more on providing room for traffic and less space for commerce, which resulted in a 'sleepy' suburb. Even though Osdorp could not provide space for smaller businesses it had issues with congested traffic (iamsterdam, n.d., B).

Slotervaart

Simultaneously with the construction of Geuzenveld, Slotervaart was built on the other side of the Sloterplas. The neighbourhood gets its name from the old waterway that once connected the 'Overtooms Huis' with the village of Sloten (vanesterenmuseum, n.d.; D).

A former agricultural area, Slotervaart was often referred to as Amsterdam's 'vegetable patch' (iamsterdam, n.d., C). In the 1950s and 60s the neighbourhood was built up as a functionalist mix of low-, medium- and high-rise apartment blocks (vanesteren, n.d., D; iamsterdam, n.d., C). Of these post-war buildings, the 'saw tooth houses' are the most well-known, while the large Sloterhof complex built in 1959 has reached 'Rijksmonument' (national monument) status (vanesterenmusem, n.d., D).
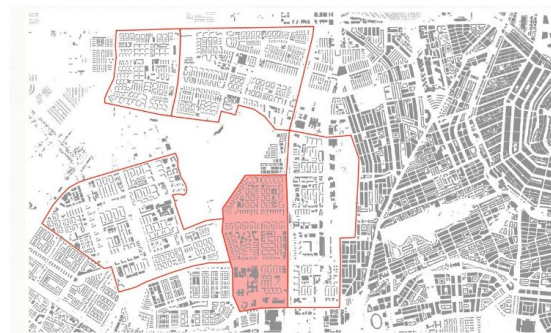
Figure 10 - Neighbourhood of Slotervaart
(source: adapted from http://vaneesterenmuseum.nl/en/garden-cities/slotervaart/)

Overtoomse Veld

In 1959 the construction of the Overtoomse Veld began. Similar to the neighbouring district Slotervaart, Overtoomse Veld contains a mix of low-, medium-, and high-rise buildings. The August Allebéplein square marks the centre of the neighbourhood.

The Overtoomse Veld has always been known for its many different functions and amenities aimed at the entire city, such as the GVB bus depot, the World Fashion Centre and hospitals (vanesterenmuseum, n.d., E).
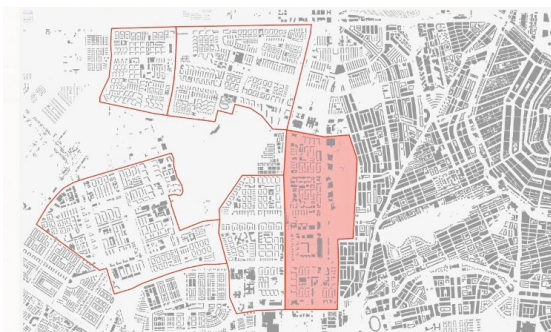
Figure 11 - Overtoomse Veld
(source: adapted from http://vaneesterenmuseum.nl/en/garden-cities/western-city-gardens/overtoomse-veld/)

These rather pleasant neighbourhood descriptions, portray a rosier reality than is backed up by facts. The truth is, the needs and requirements of citizens had changed in the decades following the completion of the post-war development. In the early 90s residents no longer saw their housing needs sufficiently reflected in the, not only monotonous, but also low quality housing stock. The public green spaces were abandoned with no one to maintain them, turning into zones which created unease and were better avoided. In consequence the socioeconomic status of the whole district was in rapid decline. Poverty, misbehaving children, school dropout rates on the rise, gangs of problematic teenagers, street criminality and in some cases even Islamic radicalisation became part of the Nieuw West's issues (van Heelsum, 2007).

Even though the safety index improved between the years 2004 and 2006, the problem characteristics remained. Large amounts of social housing, large numbers of unemployed, a high percentage of people living below the social minimum and a wave of rather new immigrants created the sketchy image of Nieuw West (Heelsum, 2007).

The multicultural population grew, so that in 2013, over half of the 133,000 residents belonged to minority groups. Additionally, the household income per year was 15-20% lower in Nieuw West than the Dutch average. Almost a quarter of the families earned minimum wage, while a high number lived off welfare (Dienst Ruimtelijke Ordening Amsterdam, 2013). The persistent social problems, unhealthy lifestyles and out-dated, underused and partly polluted open spaces had turned the Western Garden Cities into a problem concentration area, lacking social cohesion and all sense of place. The architecture of the area was seen as vital to the issues, which is why ultimately the municipality of Amsterdam drew on drastic actions.

5.1.2 Project Initiation

Already in 1993 the neighbourhoods Osdorp and Geuzenveld-Slotermeer initiated the formation of the organisation Bureau Parkstad, which was to play a primary role in the future redevelopment project. This local initiative was started to formulate future strategies for the reform of the neighbourhoods and lay down the first draft plans for the physical redevelopment (Nio et al., 2016).

This resulted in the establishment of an initial redevelopment plan in 2001 known as *Richting ParkStad 2015*. This plan was authorised by the municipality of Amsterdam and the involved housing associations and contains the ambitions of the revitalisation of the neighborhoods in Nieuw West (EUNK, 2010, Stadslente, 2016). In 2003, the first urban redevelopment plan was introduced. Together these plans laid the foundation for concrete social and spatial projects, which were approved in the course of 2005. In summary, these project plans foresaw the reduction of social housing from 41.000 apartments to 29.000, partly through the sale of these dwellings, partly through demolition. Overall, 13.300 dwellings were to be demolished, which amounts to 25% of all houses. Additionally, 24.300 new homes were to be built, densifying the area through the addition of 11.000 houses in total. This investment was estimated to cost between 4 and 5 million euro (Nio et al., 2016; Stadslente, 2016). The below table (table 6) summarises the development goals of the main stakeholders: the neighborhood initiative Bureau Parkstad, the municipality of Amsterdam and the housing associations.

Nieuw West Stakeholder Development Goals	
Stakeholder	**Development Goals**
Bureau Parkstad	a Park City (Richting Parkstad 2015): establishing stronger bonds ans connectivity between residents, create environment of stability and harmony
	overcome the stigmatised image
	increase the socio-economic position of residents
	attract new businesses
Municipality	create an urban social mix through attracting mid-income renters
	economic growth of the neighborhood
	in numbers: reduction of 12.000 social housing units, in total demolition of 13.300 dwellings, new build of 24.300 units, densification of the area by adding 11.000 units overall
	social renewal: improving educational results, social cohesion, increasing participation in society and in the labour market
Housing Associations	develop housing according to their individual mission and target group's needs

Table 6 - Stakeholder Development Goals

5.1.3. Development Process

As mentioned, the above described project is aimed at improving the Nieuw West district in three areas: social, economic and spatial. In order to achieve this goal the redevelopment takes place in form of an integrated project, following close cooperation between the municipality of Amsterdam, local city authorities, housing corporations as well as residents' organizations, welfare organizations and health insurers (EUKN, 2010).

The general ambitions of the renovation are laid out in the Richting ParkStad 2015 development plan. The Nieuw West district is further divided into 27 project areas. A separate renovation plan is being drawn up for each of these project areas in cooperation between the neighbourhoods and the housing corporations. Each renovation plan provides a substantive and financial framework for its area, including the housing programme to be realised, which social amenities will be provided, the number

of square metres of retail space to be supplied, as well as the social programmes that will be employed. Each renovation plan continuously undergoes a process of revision, becoming more tangible over time. While the housing corporations realise the residential construction, the city district is responsible for carrying out the redevelopment of public space (EUKN, 2010; Nio et al., 2016).

In 2005 the first interim evaluation of completed projects and their effects takes place. The conclusion is that despite the physical renewal, the socioeconomic position of the residents is still below the Amsterdam average. Due to these results and the fact, that current residents are not interested or able to purchase some of the renovated and new built condos, the physical redevelopment of Nieuw West stagnates. Apart from that, the project's budget is coming to an end (Nio et al., 2016; Stadslente, 2016).

Then in 2007 a change is added to the development plan. The focus is to shift to ensure the improvement of the situation of the current residents. This is to be achieved through providing more professional opportunities especially for graduates and stimulating local entrepreneurship (Stadslente, 2016).

The economic crisis of 2008 heavily impacts this undertaking. The housing corporation's works come to an abrupt hold, while the sale of new built freehold apartments stagnates. This promotes the financial issues only further, as the project funding is dependant on the sale of freehold dwellings (Nio et al., 2016).

The year 2012 is when the municipality of Amsterdam, the district Nieuw West and the housing associations together determine the project's new course. The direction points toward small-scale interventions that should be spread out over time and be less impactful (Stadslente, 2016).

However, in 2013 another obstacle is put in projects way. The investment capacity of housing corporations is reduced, meaning with the *New Housing Act 2015* (nieuwe woningwet) housing associations are not admitted to develop freehold dwellings outside the rent regulated sector anymore. This leads to the project stakeholders' joint decision to focus more on the existing stock and improving it through renovation rather than demolition (Nio et al., 2016; Stadslente, 2016).

Following the structural improvements of the housing stock, the focus will turn toward the concept of social renewal. These policies are aimed at enhancing social cohesion, improving educational results, residents' participation in society, as well as participation in the labour market. The concept of social renewal, focuses mainly on projects that are to be realized once the physical restructuring is completed. Most of these projects were established at the neighbourhood level. Examples of such projects are the virtual mini-pitches, the implementation of a social work manager for primary education, and the Moederkind Centrum, an initiative for women's empowerment. On occasion a joint approach toward social renewal was taken between the neighbourhoods. Two successful examples are the 5 O'clock Class, which is a theatre class for young people and the ten-day festival on the Sloterplas (EUKN, 2010).

5.1.4. Project Outcome

Finally, in 2016 almost 9.500 new builds were completed, while 7.000 dwellings had been demolished. This is a lot less than the originally planned 24.300 new builds, as stated in the development plan *Richting Parkstad 2015*. Also, 6.300 dwellings less than originally planned had been demolished, reflecting the impact the economic crisis in 2008 and the political reforms throughout the decades have had on the original redevelopment plans (Nio et al., 2016).

The amount of social housing was reduced from 76% in 2000 to 53% in 2016 (Nio et al., 2016; Stadslente, 2016), almost adjusting it to the Dutch average of 52% (Heelsum, 2007).

Moreover, public facilities have been added to the area, specifically educational facilities. In total 14 primary schools and 5 schools for higher education have been realized. A lot was also invested in the abandoned green spaces of de Westelijke Tuinsteden, for example the Tuinen van West, Rondje

Sloterplas, the neighbourhood parks Gerbrandy-park and Eendrachtspark, and a new park was created in Osdorp the Stadspark Osdorp. The restructuring of the main roads, with the aim of improving connectedness between new developments and refurbished ones and between the neighbourhoods, however, lacked some organizational control (Nio et al., 2016).

Reviewing these facts, the question is asked whether these interventions have led to the development aim of raising the socioeconomic level of the Nieuw West district, while improving it sociocultural and spatially as well. A common debate, which summarizes this questions is, whether Nieuw West is now a Park City (parkstad) or an Urban District (stadswijk).

Park City vs Urban District
The following arguments are based on an extensive research analysis on the effects and outcomes of the redevelopment of Nieuw West by Nio et al. (2016). The authors explored the Nieuw West district from socioeconomic, sociocultural and spatial viewpoints after the completion of the redevelopment. One of their main statements is, that Nieuw West has its own urbanity, its own power and spirit and that the outcomes of the redevelopment cannot solely be expressed in facts and socioeconomic indicators (Nio et al., 2016).

First, there are large differences to be taken into account within Nieuw West itself. The authors therefore focused on the spatial as well as sociocultural effects the physical renewal had on the postwar city, examining the improvements and meaning of the collective public spaces and the standpoints of the most important stakeholders, closely.
Based on their analysis, they are able to declare the growing heterogeneity of residents in Nieuw West as urbanisation in the socioeconomic sense. The aim of Bureau Parkstad was, however, focused on Nieuw West as a Park City not an Urban District, where the urbanisation in a social and cultural sense are more important than socioeconomically. Through looking at interventions in the public space and how a new collective mindset is formed, including the formation of new borders of private, collective and public, the authors want to determine Nieuw West's development direction. Do the improvements in public areas leave space for changes and spontaneity? According to their definition, the Bureau Parkstad's concept of a Park City depicts harmony and stability, while the Urban District leaves space for dynamics, changes and different lifestyles, resulting in more conflicting situations. Originally, it was aimed at establishing stronger bonds and a sense of collectivity between residents. On dwelling level this was partly achieved for residents sharing private inner gardens. However, in socioeconomic and sociocultural perspective Nieuw West is becoming more heterogeneous with a greater diversity of residents and larger differences in owner occupied and rented apartments (Nio et al., 2016).

It is concluded, that while the Park City concept of harmony and collectivity was the initial ambition of the redevelopment project, the Urban District reflecting diversity and conflicts was the outcome due to the social, economic and cultural dynamics found in Nieuw West.

Nieuw West Stakeholder Development Goals vs Development Outcome		
Stakeholder	Development Goals	Development Outcome
Bureau Parkstad	a Park City (Richting Parkstad 2015): establishing stronger bonds ans connectivity between residents, create environment of stability and harmony	Park City and Urban District: leaving space for dynamics, changes and different lifestyles, resulting in more conficts
	overcome the stigmatised image	the area turned into a more attractive place for families to live
	increase the socio-economic position of residents	socio-economic positions was increased for only some, creating gaps and 'pockets of poverty'
	attract new businesses	small family-run shops have opened
Municipality	create an urban social mix through attracting mid-income renters	the renewal was successful in drawing the (foreign) middle income class into the area
	economic growth of the neighborhood	on district level is may seem so, but on neighborhood level there are large discrepancies regarding the economic situation
	in numbers: reduction of 12.000 social housing units, in total demolition of 13.300 dwellings, new build of 24.300 units, densification of the area by adding 11.000 units overall	in numbers: reduction of social housing from 76% to 53% (equals the planned reduction of 12.000 social housing units), demolition of 7.000 units, new build of 9.500 dwellings
	social renewal: improving educational results, social cohesion, increasing participation in society and in the labour market	social renewal: 14 primary schools and 5 schools for higher education, investment in public areas to create space for social encounter, no connections between newly deveoped and remaining areas, generally plans for enhancing social cohesion and participation were on the sidelines or implemented late, which is why they have not caught on as expected
Housing Associations	develop housing according to their individual mission and target group's needs	policy developments made realizing their individual missions difficult, housing associations took on a more passive role

Table 7 - Stakeholder Development Goals vs Development Outcome

Evaluating the aims and outcomes
Besides this disillusioning result, there are further opinions to be heard. The Sociaal en Cultureel Planbureau stated in 2013, that all the above listed interventions have had next to no positive effect on social improvement or income level dispersion, nor on liveability and safety (Nio et al., 2016).

While the leefbaarometer of RIGO and Atlas for municipalities states that between the years 2008 and 2014 an improvement of liveability in Nieuw-West has taken place, as a consequence of the physical improvements of the housing stock. Also, the renewed public space achieved a score above average. According to their examination, Nieuw West had become a stable neighbourhood and has experienced a slightly positive development. However, in 2015 Geuzenveld-Slotermeer and Osdorp drew attention due to their overall lower score in this assessment. The average income in Nieuw West was still lower than the Amsterdam average, and large differences could be experienced between areas in which physical revitalization had taken place directly influencing the growth of income.

On top of that, the neighbourhood appreciation had risen less than expected. Surprisingly it had increased the most in Slotermeer, which experience almost no physical renewal (Nio et al., 2016).

On the other hand, the aim of Bureau Parkstadt to attract a more affluent middle class can be termed successful. Especially, the provision of housing for middle class residents from within Nieuw West has increased. This was mainly achieved through the number of new built projects, enabling the native middle class to grow and increasing the social stability in Nieuw West.

On top of that, also the foreign middle class that had grown up in Nieuw West but moved away when the neighbourhood was deteriorating is now returning. Nieuw West offers them a living environment that they know well, with already established social networks and a place where they feel at home. This particular social group feels connected to the area on a spatial and a social level (Nio et al., 2016). Generally, Nieuw West has become more attractive for residents looking for larger dwellings and more green spaces in the proximity to the city.

Yet, the sociocultural dynamics in Nieuw West are different for each neighbourhood. The mix of owner-occupied and rental apartments should have resulted in a more mixed area, which it does on a larger scale. However, when residents are asked directly, this mix is seen as a problem, hinting at a lack of collectivity. In other, more homogeneous complexes and areas the collective mindset was strengthened.

Primarily apartments without outer spaces, which were numerously developed in the early stages of revitalisation, and apartment complexes struggling to define collective spaces, are facing problems. Also, complexes consisting of a strong mix between social housing and owner occupied dwellings have been experiencing management difficulties.

The renewed parks and sports locations are used more intensively. However, the development of schools has changed little regarding the division of resident groups, and there are still the 'zwaarte scholen', which mainly problematic youngsters attend.

Nonetheless, there have also been multiple improvements, such as the refurbishment of the Sierplein, the quay of the Sloterplas in Osdorp and the realization of playgrounds, which don't get a lot of publicity but have a positive impact on the daily life of residents.

In Conclusion
On the one hand, settled resident groups in Nieuw West are now able to climb the housing ladder and newly arriving middle income households are able to reside in the district as well. The housing stock has been improved, as well as the public spaces and public facilities, such as schools. However, the physical renewal is not everything, and segregation in Nieuw West is still a hot topic.

The urbanisation that has taken place through the physical renewal, has furthered the divide in Nieuw West. The socioeconomic and demographic differences have grown on neighbourhood level. Osdorp and Geuzenveld attract many foreign residents, while Slotervaart attracts double income families (Nio et al., 2016). These differences were visible already before the redevelopment, but have since been increased. Nieuw West stays a sociocultural and socioeconomic dynamic district. Native and foreign residents have different expectations and identify differently with collective and public spaces.

Moreover, due to the demolition of many social housing dwellings, residents with low income have left Nieuw West to settle in a different area. Despite strict management, the concentration of poverty has

increased in the leftover social housing dwellings, creating so called pockets of poverty. On the district level of Nieuw West the aspired mixture has been achieved, but on neighbourhood level the differences have grown. Increasing the cohesion in such mixed areas is now the role of the streets, parks and squares, which structurally still need to be improved to enhance the lived city.

Apart from that, the focus of the redevelopment has always been functional and not on the representative meaning of architecture and public buildings for familiarity and attracting new residents. Also, commercial aspects such as small shops and leisure have not been promoted sufficiently. These things partly determine the identity of the neighbourhood and if people will settle and feel at home.

As Nieuw West is both a Park City and an Urban District it must offer spaces for dynamics, changes, urban life, meeting places, exchange and conflict, to create a place where distance and proximity are balanced in the collective concept.

All in all, this examination makes it clear, that the initial planning and implementation of the revitalisation project in Nieuw West, was focused on improving socioeconomic and physical indicators. This focus was mainly based on the analysis of indicators from the categories *people* and *property*. . Only after completion did the stakeholders foresee plans for sociocultural renewal. However, local networks and social cohesion are not eventualities to be dealt with if the timing or budget allow it. The lack of embedding indicators from the category *area* into the decision making and planning process is a fault and reflected in the bilateral redevelopment outcome.

Moreover, the scale or resolution at which a neighborhood analysis is conducted, substantially influences the results: Nieuw West on a district level shows more positive outcomes with regard to mixture and social cohesion, than results on an individual neighborhood level or even street or dwelling level. As shown, analyses in Nieuw West target the district level, as it produces more positive results, and only look at the neighbourhood level when its results add to the positive picture that wants to be painted. However, compositions on neighbourhood level contain information crucial to indicators of the *area* category. Nieuw West thus portrays an excellent example of the flaws in contemporary urban development and construction projects of the built environment to take complex local structures, represented in the indicators of the *area* category, into account.

The following chapter will turn toward an illustration of the added value of big data standards through examining the framework's performance within the above introduced context of the case study in Nieuw West.

6. The Framework's Performance

6.1. The Illustration Procedure

The aim of this illustration is to determine, whether it is possible to reproduce the development outcome as described above with the formulated blueprint. Should it be able to recreate what has been established through Nio et al.'s (2016) research, its performance is a success and it can then be assessed, where the framework could possibly be of added value in future planning and decision making processes of urban development and building construction projects.

The illustration is structured into two parts, which in the end will be brought together to complete this example and to arrive at the integrated framework, toward which the previous chapters have been leading up to.

The first part of the illustration concerns the extraction of information on big data indicators such as *Growth of Income*, *Rental Housing Supply*, *Increase in Rental Prices*, *Neighborhood Characteristics*, *Vitality and Diversity* and *Resident Displacement* from the data source Insideairbnb. As the data on Airbnb rentals in Amsterdam Nieuw West has been thoroughly prepared for research and is openly accessible through the platform Insideairbnb, actual information on the big data indicators can be extracted.

The second part of the illustration involves the derivation of information on big data indicators such as *Key Location*, *Attractiveness*, *Social Cohesion* and *Usage of the Urban Landscape* from the data sources Twitter, Flickr and Foursquare. In contrast to part one of the illustration, data required from these platforms is unstructured and is not ready to be used by researchers. As the pre-process of the information from these data sources needs a large amount of time and meticulous expertise, which lies outside the scope of this research, no actual information on these big data indicators can be derived. However, the below chapter concerning this issue, contains all necessary preparatory work to ensure the successful extraction of data in the future.

6.2. Part 1: Insideairbnb Data

Growth of Income
In order to discover whether the Nieuw West district is prone to experiencing growing household incomes and thus, experiencing a boost in the socio-economic status, the average monthly incomes from Airbnb rentals on insideairbnb.com are examined and compared to data on median monthly rents. That way, it can be concluded whether residents are achieving a surplus with regard to their finances, by renting out their apartments for short-stay through Airbnb. The data is individually available for the neighbourhoods Geuzenveld-Slotermeer and Osdorp, while Slotervaart and Overtoomse Veld are treated as one neighbourhood.

As table 8 below indicates, the average monthly income from Airbnb rentals in Nieuw West is lowest in Osdorp with 622€. While it is 30€ more in Slotervaart/Overtoomse Veld and almost 100€ per month more in Geuzenveld-Slotermeer with 747€. On average the monthly income through Airbnb in Nieuw West lies at 673€ (dated 9th April 2018).

The median monthly rents are determined by reviewing the top 20 apartments listed on Pararius on the 9th April 2018. An average monthly rental price is determined by analysing the average square meter price and size of apartments in Nieuw West, resulting in an overall median monthly rent of 1.537€ for units within the non-regulated rental sector (Pararius, 2018).

Units let within the regulated rental sector are capped at 711€ per month (Gemeente Amsterdam, n.d.).

1. Determine Big Data Indicator	2. Define the Data Source	3. Collect Data		4. Analyse Data	5. Data Output
Growth of Income	Airbnb Data	Geuzenveld-Sloter meer	average monthly income: **747€**	average income per month: **673€**	
		Osdorp	average monthly income: **622€**	average montly rent (free market): **1.537€**	
		Slotervaart + Overtoomse Veld	average monthly income: **652€**	average monthly rent (regulated market): **711€**	

Table 8 - Growth of Income identified through Airbnb Data and Advertised Rentals on Pararius

Rental Housing Supply

In order to receive an indication on the supply of rental housing, data on the number of Airbnb rentals and the typology (entire home or apartment, private room or shared room) is collected for each neighbourhood in Nieuw West from insideairbnb.com.

The amount of entire homes or apartments for rent is the essential information for this indicator. With 356 Airbnb rentals in total, Slotervaart/Overtoomse Veld together have more than double the amount of rentals compared to Osdorp and Geuzenveld-Slotermeer. They also have the highest percentage of entire homes or apartments for rent (64%). To set the individual neighbourhood data into perspective it is compared to the Amsterdam average amount of entire homes or apartments as well as the neighbourhood with the highest and the neighbourhood with the lowest amount of this typology for rent (see 4. Analyse Data in table 9).

Finally, an average is created for Nieuw West, resulting in 55,9% of Airbnb rentals being entire homes and apartments. Compared to the Amsterdam average the amount of entire homes and apartments for rent in Nieuw West is over 20% lower, but compared to the area with the lowest amount of this typology, Nieuw West has 15% more entire homes taken off the market.

1. Determine Big Data Indicator	2. Define the Data Source	3. Collect Data					4. Analyse Data		5. Data Output
Rental Housing Supply	Airbnb Data	Geuzenveld-Sloter meer	total: 198	entire home/apt.: 112 (56,6%)	private room: 84	shared room: 2	in Nieuw West on average 55,9% of rentals are entire homes and apartmens	Amsterdam average: 79,5%	
		Osdorp	total: 140	entire home/apt.: 66 (47,1%)	private room: 72	shared room: 2		Westerpark (highest) 87,1%	
		Slotervaart + Overtoomse Veld	total: 356	entire home/apt.: 228 (64%)	private room: 127	shared room: 1		Gaasperdam-Driemond (lowest) 41%	

Table 9 - Rental Housing Supply identified through Airbnb Data

Increase in Rental Prices

Whether an area experiences a growth in rental prices, can be determined by looking at the number of Airbnb rentals within an area and the nightly prices. The same conditions as above apply to the collection of this information regarding the accessibility and division of neighbourhoods (see table 10 below).

Generally it is believed that the higher the nightly price of an area, the higher in return the monthly rental price of common apartments and houses.

As stated above, Slotervaart/Overtoomse Veld count the most Airbnb rentals overall. The average nightly price is, however the same in Slotervaart/Overtoomse Veld and Geuzenveld-Slotermeer and only 5€ lower in Osdorp. The average for Nieuw West results in 103,5€ per night. Again, this data is compared to the overall Amsterdam average, the neighbourhood with the highest nightly prices and the neighbourhood with the lowest nightly prices. It is observed, that Nieuw West's nightly prices are about 40€ below the Amsterdam average and looking at the highest and lowest nightly prices, fees in Nieuw West sit at the lower end of the scale.

1. Determine Big Data Indicator	2. Define the Data Source	3. Collect Data			4. Analyse Data		5. Data Output
Increase in Rental Prices	Airbnb Data	Geuzenveld-Sloter meer	total: **198**	average nightly price: **111€**	in Nieuw West on average one night costs **103,50€**	Amsterdam average **142€**	
		Osdorp	total: **140**	average nightly price: **96€**		Centrum-West (highest) **180€**	
		Slotervaart + Overtoomse Veld	total: **356**	average nightly price: **111€**		Gaasperdam-Driemond (lowest) **85€**	

Table 10 - Growth in Rental Prices identified through Airbnb Data

Neighborhood Characteristics

In order to uncover certain characteristics of neighbourhoods the amount of Airbnb rentals in that area can give an indication. Large amounts of rentals are associated with larger number of tourists, which can create tension, with regard to aspects such as noise, congestion or parking issues. Correlating the number of Airbnb rentals with information on socio-economic characteristics from census data or by reviewing regulations with regard to tourism activities, can support such claims.

The number of Airbnb rentals in the neighbourhoods of Nieuw West (694) is distinctly lower compared to the number of rentals in the areas most densely populated with Airbnbs: Centrum-West (2.175), de Pijp (2.395) or Oud-West (3.311).

Reviewing regulations regarding tourism activities in Nieuw West or generally in Amsterdam, the discovery is made, that these are quite strict and will continue to regulate activities, such as the letting of units through Airbnb, even further (Coldwell, 2014). From January 2019 onward, the maximum amount of days for residents to let their homes through Airbnb will be reduced to 30 days a year.

1. Determine Big Data Indicator	2. Define the Data Source	3. Collect Data		4. Analyse Data		5. Data Output
Neighborhood Characteristics	Airbnb Data	Geuzenveld-Sloter meer	total: **198**	in Nieuw West in total **694** Airbnb rentals	Centrum-West: **2.175**	
		Osdorp	total: **140**		De Pijp: **2.395**	
		Slotervaart + Overtoomse Veld	total: **356**		Oud-West: **3.311**	

Table 11 - Neighbourhood Characteristics identified through Airbnb Data

Vitality and Diversity of a Neighbourhood

To establish the vitality and diversity of a neighbourhood, two sets of information are required. The first concerns the mapping of Airbnb rentals and hotel locations, as an indicator for the liveliness of that area with data from insideairbnb.com and data.amsterdam.nl. The second compares the Airbnb host origin and spoken languages from insideairbnb.com to the people registered through census data speaking that language, to determine the diversification of the neighbourhood.

For the indicator vitality it is believed that in areas with Airbnb activity, but no or next to no hotel activity, Airbnb functions as an 'area booster', adding to the liveliness of the place. Looking at the map below (Figure 11), containing the locations of Airbnb rentals throughout Amsterdam (red and green dots), as well as the locations of all Amsterdam hotels (blue dots), it becomes apparent that the centre of Amsterdam has a high activity both for Airbnbs and hotels. Zooming in on the case study area Nieuw West, it holds, as we have established 694 Airbnb rentals in total, compared to the number of five hotels in the district.
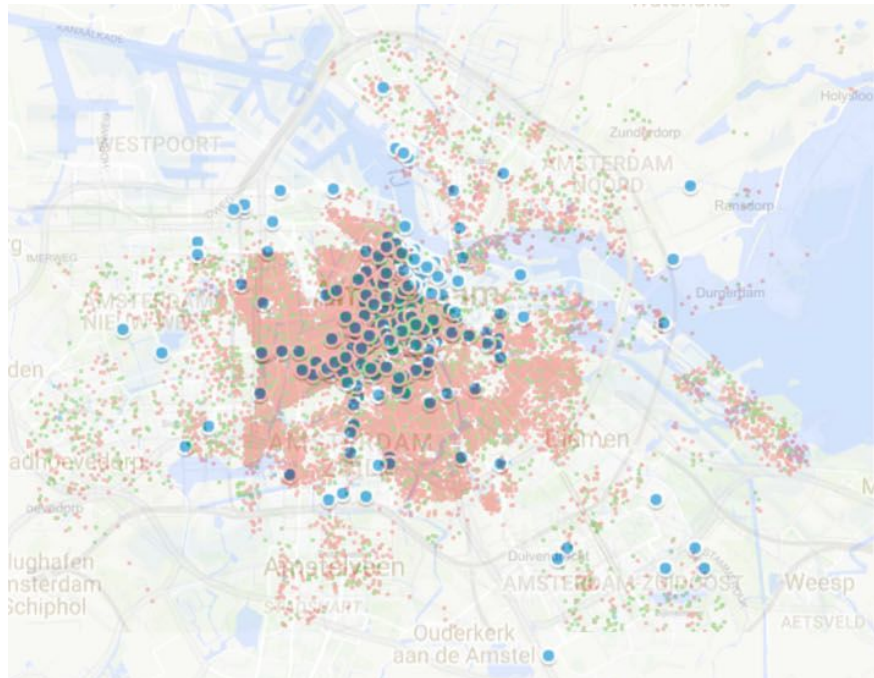
Figure 12 - Neighbourhood Vitality identified through Airbnb Data and Amsterdam City Data

The expected diversity of the neighbourhood from the multilingual Airbnb host profiles from insideairbnb.com is examined in comparison to census data reflecting residents' origins (Table 12). For this demonstration the host profiles of the top 20 hosts per neighbourhood have been examined.

Naturally, English is the primary language of Airbnb, followed by Spanish, German, Italian and French. All five are found repeatedly within the three case study neighbourhoods. Portuguese is only spoken by hosts in Geuzenveld-Slotermeer and Slotervaart/Overtoomse Veld. Turkish and Chinese are respectively spoken by only one of the top hosts in Geuzenveld-Slotermeer. Greek is listed by just one host in Osdorp, while Magyar is spoken by a host in Slotervaart/Overtoomse Veld.

While Amsterdam as a whole is 49,5% Dutch and 50,5% foreign ancestry, apparently non-Westerners are concentrated in Nieuw West (Worldpopularionreview, 2018). Identifying the exact origins of the foreign population in Nieuw West, even more specifically per each neighborhood of Nieuw West requires more in depth statistical analysis than can be offered at this time.

1. Determine Big Data Indicator	2. Define the Data Source	3. Data Collection		4. Analyse Data	5. Data Output
Diversity	Airbnb Data	Geuzenveld-Slotermeer	English, Spanish, Italian, German French, Portugese, Chinese, Turkish	most commonly spoken: **Western-European languages** (Spanish, Italian, French, German and English)	
		Osdorp	English, German, French, Italian, Spanish, Greek	at times: **Portugese**	
		Slotervaart + Overtoomse Veld	English, Spanish, Portugese, Italian, German, French, Magyar	rarely: **Chinese, Turkish, Greek, Magyar**	

Table 12 - Neighbourhood Diversity identified through Airbnb Data and Census Data

Resident Displacement

Whether a neighbourhood is exposed to resident displacement can be read from the number of Airbnb rentals in an area and the typology of these rentals (entire home or apartment, private room or shared room). The higher the amount of entire homes and apartments for rent the more likely the

displacement of residents takes place, as housing is made inaccessible through the preferred short-stay renting offered through Airbnb.

The below table (table 13) displays the same data as required for the indicator *Rental Housing Supply.* Only this time, the average amount of entire homes and apartments for rent in Nieuw West is compared to the migration rate listed in census data.

From the website statline.cbs.nl the exact number of people migrating from and arriving in Zuid-Holland in the year 2016 could be determined. While, 29.860 individuals migrated from Zuid-Holland, 48.510 individuals arrived in Zuid-Holland within the same year. Overall, this means the population in Zuid-Holland is growing.

1. Determine Big Data Indicator	2. Define the Data Source	3. Collect Data			4. Analyse Data	5. Data Output
Resident Displacement	Airbnb Data	Geuzenveld-Sloter meer	total: **198**	entire home/apt.: **112 (56,6%)**	in Nieuw West in total **694** Airbnb rentals of which **406** entire home/apartment	
		Osdorp	total: **140**	entire home/apt.: **66 (47,1%)**		
		Slotervaart + Overtoomse Veld	total: **356**	entire home/apt.: **228 (64%)**		

Table 13 - Resident Displacement identified through Airbnb Data and CBS Statistics Netherlands

6.3. Part 2: Social Media Data

Key Locations

Identifying the top tweeted-from locations can ultimately reveal how certain spaces in the neighborhood are used. In order to do so, the tweet activity of individuals in Nieuw West is followed to determine locations they tweet from on a regular basis. By identifying the top tweeted-from location for an individual and correlating this data with the time of day at which the user posts his/her tweets, can reveal the function of that location, e.g. home, work or the favourite cafe (Table X). If the most tweeted-from location lies outside Nieuw West, the individual's activity is not valuable for this research.

1. Determine Big Data Indicator	2. Define the Data Source	3. Collect Data	4. Analyse Data	5. Data Output
Key Locations	Twitter	# of users' tweets at certain locations	correlate users' top tweeted-from location with the time of day	

Table 14 - Key Locations tracked through Individual's Twitter Activity

Attractiveness

The level of attractiveness of an area can be, firstly determined through the amount of Twitter messages in Nieuw West, which determined the popularity of the place, compared to the amount of tweets in a known popular place such as the city centre of Amsterdam. Secondly, this can be identified by examining the share of Twitter messages with additional content compared to the overall amount of tweets in Nieuw West, which reveals the activity in an area. Correlating this information on popularity and activity results in the level of attractiveness of Nieuw West compared to Amsterdam's city centre. Naturally, this analysis can also be broken down per neighbourhood in Nieuw West, to compare the attractiveness between each of these. Additionally, the attractiveness of Nieuw West can be analyzed by reviewing the share of photographs posted on Flickr containing the topic 'Nieuw West' and comparing the number of photographs within this topic to the number of images of other topics in the vicinity.

1. Determine Big Data Indicator	2. Define the Data Source	3. Collect Data	4. Analyse Data	5. Data Output
Attractiveness	Twitter	# tweets in an area + # tweets with added content	correlate the shares of popularity and activity	
	Flickr	# photographs containing the topic	compare to other topics' amount of photographs in the vicinity	

Table 15 - Attractiveness identified through Twitter and Flickr data

Social Cohesion

Determining the level of social cohesion in an area requires the establishment of the social Twitter network of users in Nieuw West, the placenetwork of users and the types of venues within the placenetwork from Foursquare. By correlating data on the placenetwork, the venue type and location bridging and bonding places are revealed. Correlating the data on users' Twitter social network with deprivation areas identified through a study by the University College London from 2012 called 'The INEQ-CITIES Atlas' on socially and economically deprived areas in Amsterdam, based on statistical data from CBS Nederland, the urban social mix can be determined. Correlating the results from both examinations reveals the level of social cohesion within Nieuw West.

1. Determine Big Data Indicator	2. Define the Data Source	3. Collect Data	4. Analyse Data	5. Data Output
Social Cohesion	Twitter	social network (who follows who) and placenetwork (who goes where) of users	I. Correlate: placenetwork, venue type and location	
	Foursquare	venue types within the placenetwork	II. Correlate: social network and deprivation areas	

Table 16 - Social Cohesion identified through Twitter and Foursquare data

Usage of Urban Landscape

Determining how the urban landscape in Nieuw West is used in a wider sense than the key locations, requires the derivation of all check-in data that can be retrieved from the Twitter public timeline. Correlating this data with the venue information for all types of locations in Nieuw West from the Foursquare API reveals the characteristics of Nieuw West as a district, as well as of its users, by determining clusters of similar types of venues in close proximity with each other. This analysis can also be applied on the neighborhood level for higher resolution results.

1. Determine Big Data Indicator	2. Define the Data Source	3. Collect Data	4. Analyse Data	5. Data Output
Usage of Urban Landscape	Twitter	number of check-ins shared on Twitter	clustering similar venues in proximity and correlating this information with the amount of check-ins	
	Foursquare	venue types within the area		

Table 17 - Usage of Urban Landscape identified through Twitter and Foursquare data

6.4. Illustration Outcome

The results of part one of the illustration are summarised in the table below (table 18) and elaborated on in the following. As part two of the illustration took place without actual data, no results for Nieuw West can be formulated on this behalf.

Results	
Growth of Income	Slight growth of income as Airbnb rentals create a monthly financial support for their hosts
Rental Housing Supply	Negative impact is trivial
Increase in Rental Prices	Negative impact is trivial
Neighborhood Characteristics	Negative impact is trivial
Vitality and Diversity	Boost in vitality confirmed; Diversity confirmed but extent to which and exact mix still open
Resident Displacement	Negative impact is trivial
Key Locations	-
Attractiveness	-
Social Cohesion	-
Usage of Urban Landscape	-

Table 18 - Results of the Demonstration

Growth of Income
The average monthly income from Airbnb does not cover the free market rents. However, it comes very close to the regulated rent. Therefore it is concluded, that on average the income from Airbnb is a financial support. The extent to which varies, depending on the economic situation per household and whether that household lives in a regulated or non-regulated rental unit. For households residing in rent-capped apartments the average income comes very close to covering the monthly rent. On the other hand, literature states, that Airbnb is usually operated by middle-income households (Dudas et al., 2017; Quattrone et al., 2016). As the average income from Airbnb is comparable to receiving rent from a rent-regulated property, it is considered an added value and therefore an addition to the average income.

Rental Housing Supply
While more than half of the Airbnb rentals offered in Nieuw West are entire homes or apartments and thus, frequently taken off the housing market entirely, the overall number of rentals is rather low. Therefore, the potential pressure created on the rental housing supply is considered trivial. Nevertheless, the fact that over 50% of rentals are frequently completely relinquished to visitors could lead to perceptible consequences in the future.

Increase in Rental Prices
As the average nightly fees for Airbnb rentals in Nieuw West are at the lower end of the scale, it can be concluded, that a growth in rental prices for common apartments will unlikely be influenced by Airbnb rental prices.

Neighborhood Characteristics

Due to the fact that the number of Airbnb rentals is rather low in Nieuw West and also, that Airbnb and tourism activities are strictly regulated, the negative impacts on neighbourhood characteristics such as noise, congestion or parking issues will turn out rather ineffective.

Vitality and Diversity

As Nieuw West holds 694 Airbnb rentals in total, while there are only five hotels, it can be concluded that the area is experiencing a positive push with regarding its vitality.

The fact that Nieuw West is supposedly known as a place attracting non-Western immigrants, while most foreign languages spoken by Airbnb hosts are Western foreign, is somewhat contradictory. On the one hand, these results point toward Airbnb being a platform used mainly by Westerners and therefore represent mainly Western foreign hosts. On the other hand, the prejudice, Nieuw West is mostly populated by non-Western foreigners could be wrong. This impression is also in line with the study results by Nio et al. (2016), who have confirmed an increase in Western, especially, Western European immigrants in Nieuw West Either way the, an indication toward the diversity of the area is given, the level of diversity and diverse exactly how, is another question.

On the other hand, without confirming this observation through census data, it could also be a confirmation of the assumption postulated by Dudas et al. (2017) and Quattrone et al. (2016), that Airbnb is limited in access to white, middle-income Westerners.

Resident Displacement

The overall growing population in Amsterdam, combined with the fact of rather low amount of Airbnb rentals in Nieuw West, speaks for low resident displacement through Airbnb. However, displacement through demolition of social housing or through attracting the middle-income class can still take place, but would not be represented within this analysis.

From this illustration it becomes apparent, that the slight number of Airbnb rentals in Nieuw West, compared to e.g. Centrum-West, is so inferior, that the impact on some indicators is trivial. Leading to the matter, that no conclusive statement can be made with regard to some indicators, such as *Rental Housing Supply, Increase in Rental Prices, Neighborhood Characteristics* and *Resident Displacemen*t, other than that they do not seem to be negatively affected.

In order to substantiate the illustration outcomes they are compared to the development outcomes found by Nio et al. (2017) in the following.

6.4.1. Substantiating the Results

In table 19 below, the development outcomes are grouped into six topics: *I. Park City or Urban District, II. Attractive for Families and (foreign) middle income households, III.Strengthened Socio-economic position vs 'pockets of poverty', IV. Strengthened retail and business sector, V. Social Cohesion* and *VI. Adequate Housing Supply*. Each of these topics reflects the original development goals of all three stakeholders. The middle column then depicts the development outcomes per topic as introduced in chapter 5.1.4., while the results retrieved through the big data standards with regard to these topics are detailed on the far right. This direct comparison of development outcome and results of the framework's illustration aims at determining its successful performance. More details on this comparison are outlined below.

Development Outcome vs Framework Performance		
	Development Outcome	Results Demonstration
I. Park City or Urban District	Park City and Urban District: leaving space for dynamics, changes and different lifestyles, resulting in more conficts	Urban District
II. Attractive for families and (foreign) middle-income households	the area turned into a more attractive place for families to live; the renewal was successful in drawing the (foreign) middle income class into the area	More dynamic, diverse and supply of housing makes the place attractive for families and mid-income
III. Strengthened Socio-economic positions vs 'pockets of poverty'	socio-economic positions was increased for only some, creating gaps and 'pockets of poverty'; on district level is may seem so, but on neighborhood level there are large discrepancies regarding the economic situation	added income from Airbnb only for some
IV. Strengthened retail and business sector	small family-run shops have opened	-
V. Social Cohesion	social renewal: 14 primary schools and 5 schools for higher education, investment in public areas to create space for social encounter, no connections between newly deveoped and remaining areas, generally plans for enhancing social cohesion and participation were on the sidelines or implemented late, which is why they have not caught on as expected	-
VI. Adequate Housing Supply	policy developments made realizing their individual missions difficult, housing associations took on a more passive role	While no displacemet through Airbnb, through demolition still possible

Table 19 - Development Outcome vs Framework Performance

I. Park City or Urban District

According to Nio et al. (2017) the development has turned out to be a hybrid between a Park City and an Urban District. Leaving space for dynamics and changing lifestyles within an harmonious environment. The illustration of the framework on the other hand points more toward an Urban District as the area is characterised by its diversity, which could potentially lead to more conflicting situations, and the vitality boost Nieuw West is experiencing through the attraction of visitors.

II. Attractive for Families and (foreign) middle income households

Nio et al. (2017) have found that the area turned into a more attractive place for families to live. Moreover, they point out the renewal's success in terms of drawing the (foreign) middle income class into the area. Also the illustration reflects this position. The vitality boost makes Nieuw West more dynamic and thus, more attractive for families. Additionally, the fact that the area is open to change and diverse and enables the strengthening of socio-economic conditions for some, corresponds to the development outcome.

III. Strengthened socio-economic position vs 'pockets of poverty'

On the downside, the authors have concluded, that socio-economic positions were increased for only some residents, creating gaps and 'pockets of poverty'. Especially, when examining the neighborhood level of Nieuw West, large discrepancies regarding the economic situation are found in the population. Likewise, the illustration through the framewor revealed that the added income was restricted to few, but it could not be determined to whom. The presumption remains, that Airbnb is a platform used by the middle income households, which leads to the conclusion that it is them financially benefiting from Airbnb, creating greater gaps of socio-economic status.

IV. Strengthened Retail and Business Sector

No data.

V. Social Cohesion

No data.

VI. Adequate Housing Supply

The fact, that policy developments made it difficult for housing associations to realize their individual missions, could be part of the reasons why the development is criticized today. Due to their passive role, it was impossible for them to enforce the equal development of housing for each target group, also resulting in the creation of 'pockets of poverty', as low income households are forced to cluster around the remaining accommodations they can afford. The illustration does not reflect displacement through Airbnb directly, however the analysis is not able to reflect possible displacement through demolition or raised rents in the aftermath of reconstruction either.

Except for the tendency to lean toward an Urban District, the framework, despite its previously mentioned limitations, and the development outcome reflect the same effects the redevelopment had on Nieuw West. Hence, the aim of this illustration to examine, whether the framework is capable of reproducing the development outcome as described, is successful. However, it is important to keep in mind, that this could also be the case, due to the fact, that only a limited amount data was retrieved and perhaps therefore, contradictory results could not be produced in the first place. In a more data-rich environment, such as e.g. Centrum-West, it is believed, that such an illustration would be more fruitful. Nonetheless, the overall performance of big data standards, taking into account said limitations, in form a framework is successful in providing more detailed and real-time data to the information collection.

The following chapter, thus introduces the integrated framework and proposes, where it can possibly be of added value in future planning and decision making processes of urban development and building construction projects.

6.5. The Integrated Framework

Urban development and building construction processes undergo multiple phases of development. The below scheme (figure 12) depicts the rational urban planning process according to Berke et al. (2006) (in Yigitcanlar&Teriman, 2013). The above illustration of the introduced big data standards has taken place at the end of such a process within the evaluation phase, where the implemented plan is monitored and amended. This was deemed necessary to examine the frameworks ability to perform successfully.
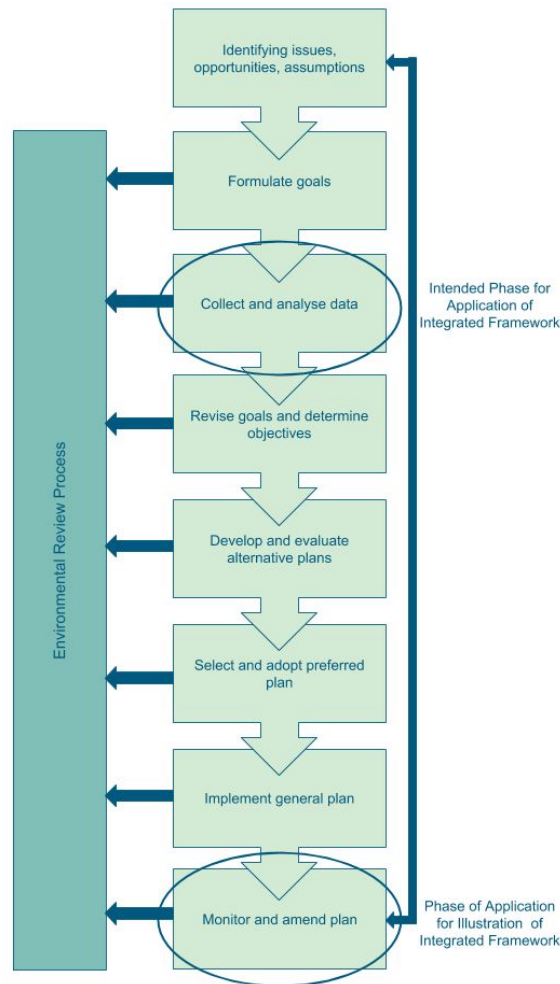
Figure 13 - Rational Urban Planning Process
(adapted from Berke et al., 2006 in Ygitcanlar&Teriman, 2013)

However, the original aim was to develop a blueprint of information collection through big data useful for the planning and decision making phase, where data is collected and analysed prior to determining the objectives of the development plan. This should help to prevent the occurence of negative consequences of contemporary urban development and construction projects in the built environment. Regardless of the phase in which it is applied, the workings of the developed framework stay the same. It is used to analyze urban dynamics on a high resolution and in real time. In the planning and decision making phase this is done before any development has taken place, in order to establish a holistic neighborhood profile on which educated decisions can be based. To ensure the comprehensiveness of information, the big data standards are complemented with traditional means for information collection on the neighborhood level (table 20).

The Integrated Framework					
	1. Determine Indicator	2. Define the Data Source	3. Collect Data	4. Analyse Data	5. Data Output
Big Data Indicator	Growth of Income	Airbnb Data	average monthly income from Airbnb rentals within a specified area	compare average monthly income from Airbnb to average monthly rents (regulated and non-regulated sector)	-
	Rental Housing Supply	Airbnb Data	total number of Airbnb rentals within a specified area and typology of rentals	determine the amount of rentals per typology: entire home/apartment, private room and shared room	-
	Increase in Rental Prices	Airbnb Data	total number of Airbnb rentals within a specified area and average nightly price	compare average nightly prices within specific area to maximum and minimum prices in neighboring areas	-
	Neighborhood Characteristics	Airbnb Data	total number of Airbnb rentals within a specified area	compare total number of rental within specific area to number of rentals in neighboring areas	-
	Vitality and Diversity	Airbnb Data	Airbnb host languages and total number of Airbnb rentals and other tourism accommodations, e.g. hotels within a specified area	compare most commonly spoken host languages to individuals registered in the area originating from a country speaking that language / overlap total number of Airbnb rentals and other tourism accommodations	-
	Resident Displacement	Airbnb Data	total number of Airbnb rentals within a specified area and number of rentals with typology entire home/apartment	compare the number of entire homes/apartments to the total number of Airbnb rentals	-
	Key Locations	Twitter	# of users' tweets at certain locations	correlate users' top tweeted-from location with the time of day	-
	Attractiveness	Twitter	# tweets in an area + # tweets with added content	correlate the shares of popularity and activity	-
		Flickr	# photographs containing the topic	compare to other topics' amount of photographs in the vicinity	
	Social Cohesion	Twitter	social network (who follows who) and placenetwork (who goes where) of users	I. Correlate: placenetwork, venue type and location II. Correlate: social network and deprivation areas	-
		Foursquare	venue types within the placenetwork		
	Usage of Urban Landscape	Twitter	number of check-ins shared on Twitter	clustering similar venues in proximity and correlating this information with the amount of check-ins	-
		Foursquare	venue types within the area		
Traditional Indicator	Median Income	Employment organisations, tax department and Dutch central bank	average monthly income of a population group within a specified area	compare to larger scale or other specified areas	-
	Education Level	Employment organisations and municipal registry	years spent on school and higher level education by a population group within a specified area	evaluate the discrepancies within the population group	-
	Education Status	Municipal registry	current state of an individual's enrollment within an educational institution	compare the number of individuals enrolled to the whole population group	-
	Employment Status	Tax department and municipal registry	current state of an individual's employment	compare the number of individuals employed to the whole population group	-
	Elderly	Municipal registry	age and health condition of an individual	categorize individuals on a scale young and old, healthy and not healthy	-
	Origin	Municipal registry	ancestry, specifically the country of origin of an individual	map the origins of individuals within the overall population group that are divergent of the majority's origin	-
	Population Growth	Municipal registry	population development over time within a specified area	compare composition of population at various moments in time	-
	Marital Status	Tax department and municipal registry	family position of an individual	categorize individuals according to married, single, divorced, widowed, etc.	-
	Tenure	Municipal registry and land registry	tenureship status of a property	categorize property according to rented out, owner occupied, vacant etc.	-
	Age	Land registry	years that have passed since the original construction or since the most recent renovation of a property	categorize property according to its final construction/last renovation	-
	Conract Rent	Statistical real estate reports compiled by real estate firms and housing corporations registry	rental amount to be paid by the tenant on a monthly, quarterly or yearly basis	compare property's rental amount to other properties within the specified area	-
	Average House Price	Statistical real estate reports and municipal registry	median amount for which a house is sold within an area	compare property's sales price to other properties within the specified area	-
	HIST	Land registry	proximity to historic sites	map the number of historic sites within the specified area	-

Table 20 - The Integrated Framework

The integrated framework foresees the compilation of a complete neighbourhood profile to support the better management of city dynamics in the context of urban development and construction projects of the built environment. This is achieved through the information collection on priorly defined indicators (step 1). In addition to the traditional indicators, usually deployed for this purpose, big data indicators have been identified, as they allow for real-time information to be retrieved at a higher resolution. For each indicator one or more data sources have been determined from which the required information can be collected (step 2). Precisely which information is necessary to be gathered, is defined in the data parameter in step 3. Finally, the data is analysed, through comparison, correlation or combination of information (step 4). Before in step 5 a data output is created, which can ultimately be

cross-referenced with other data for valid results. The combination of all data outputs of all indicators compiled, can provide conclusive insights into city dynamics on neighborhood level.

In the specific case of Nieuw West the application of the integrated framework during the planning and decision making phase, could have prevented the bilateral development outcome. By retrieving detailed information on neighborhood level, especially on indicators such as *growth of income*, *increase in rental price*s, *vitality and diversity* and naturally, *social cohesion* and *usage of urban landscape*, the heterogeneity of the area could have been revealed and strategies to target the large differences in socio-economic status and deprived spaces could have been formulated on a higher resolution than district level, providing indispensable awareness for local challenges.

Not only in the case of Nieuw West, but overall, the integrated framework is thought to be the first step toward a future instrument helpful during the planning and decision making phase of urban development and building construction projects. Assessing the status quo of the neighborhood which will be impacted by such developments, is vital to tailoring the project to the neighborhood's specific opportunities and needs. Having a complete profile of the area at hand would enable decision makers and planners to base their choices on researched facts, rather than on superficial or outdated knowledge and comparable projects from practice.

Any individual or party involved in planning and decision making process of urban development and building construction projects is eligible and even encouraged to apply such a future instrument to their specific project context. Generally, these individuals or parties would be municipalities, urban planners and developers, but also resident organisations or neighborhood initiatives could already make use of an integrated framework as depicted above, provided they have the means and expertise to collect and analyse the data. It is believed, that the integrated framework forms the basis of future information collection and is particularly able to give voice to more passive stakeholders during the planning and decision making phase, by paving the way toward retrieving detailed neighborhood information in other ways than through traditional census data. The following chapter therefore examines the changing role of housing associations in the context of the redevelopment project in Nieuw West and whether such an integrated framework could empower housing associations in future planning and decision making processes, despite their recently more passive role.

7. Practical Implications of the Integrated Framework

7.1. The Role of Housing Associations

Throughout the above described development process, the involved actor's roles and responsibilities have shifted on multiple occasions due to external circumstances influencing the financial situation and organisational setting of the stakeholders (EUKN, 2010). As some of these events have been vital to the way in which this decade-long development process has turned out, the changing actor roles and responsibilities deserve a closer look. Most importantly, the changing roles and shift in responsibilities the housing corporations involved in the redevelopment of Nieuw West have experienced, are examined more closely, due to their particular position: the amount of ownership of housing corporations is high in Nieuw West and lies at above 60% (Heelsum, 2007). At the same time, ownership is spread over nine different organizations.

In the following, the policy developments with regard to the regulation of housing corporations in the Netherlands will be introduced and connections will be depicted with regard to the impact these developments have had on the revitalisation of Nieuw West.
To begin, the nine housing corporations throughout Nieuw West are introduced.

7.1.2. Housing Associations in Nieuw West
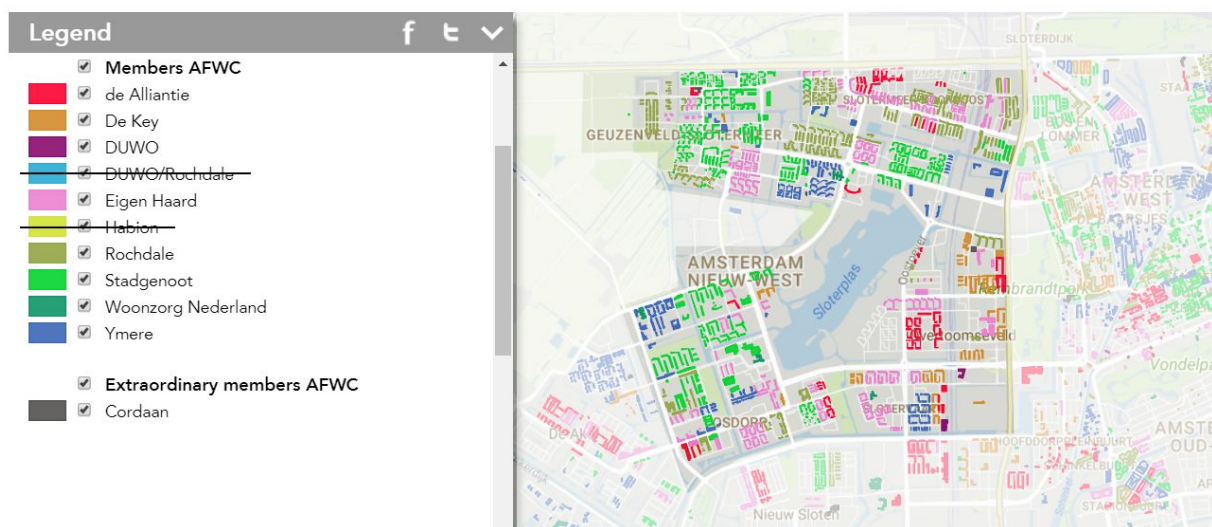
Figure 14 - Housing Association property diversificatio
(source: adapted from https://maps.amsterdam.nl/afwc/?LANG=en)

Figure 13 depict the diversification of housing associations property in Nieuw West. The Coloured properties indicate partial or full ownership by a housing corporation. The diversification of ownership becomes apparent through the number of differently coloured properties. Below a brief profile is given of each organisation to portray their well defined visions and missions.

de Alliantie
De Alliantie is a housing corporation with the aim to provide affordable housing for people with a modest income. They rent, maintain build and sell 53.000 dwellings, rentals and dwellings for owner occupancy, in the regions Amsterdam, Almere, Amersfoort and Gooi en Vechtstreek. As an approved organisation, in compliance with the Housing Act, they stand under the administration of the Housing Corporation's Authority (Autoriteit Woningcorporaties).

Since 2001, they have undergone multiple mergers with different corporations, all six of which now carry the name *de Alliantie* (de Alliantie, n.d.).

de Key

Under the guideline space for development ('Ruimte voor beweging') the housing association *de Key* focuses on supporting people entering the housing market. Especially young people with limited means have difficulty setting foot in the housing market of Amsterdam. *De Key* aims at balancing the overpriced inner city market by offering affordable social housing in that area. Together with other housing corporations they intend to offer dwellings to all groups eligible for social housing. In order to support the city as a whole they also deal with social issues by asking what it is that Amsterdam needs. This is how they arrived at today's direction of focusing on people entering the housing market (de Key, n.d.).

DUWO

Started in Delft after the Second World War, DUWO has grown to be the largest student housing association in the Netherlands. It provides approximately 30,000 homes in the regions of Amsterdam, Delft, The Hague, Leiden as well as in Deventer.

While it continues to be difficult for students to access affordable and at the same time good quality housing, DUWO continues to play a leading role in reducing the housing shortage and further improves its service. In 2013, DUWO became an online student housing corporation,making it possible for students to take care of their affairs online. From early 2015, also international students are able to use the online services. Providing accommodation to as many Dutch and foreign students as possible through affordable housing that meets all modern demands is what DUWO is all about (DUWO, n.d.).

Eigenhaard

Eigenhaard have made it their cause to accommodate as many people as possible in affordable housing of good quality and in beautiful neighbourhoods.

As a housing corporation they believe that living crosses the borders of the own home. In order to guarantee the best possible communal living environment a service attuned to specific customer needs is key, while balancing social, spatial and economic matters. Building sustainable relationships with their partners, as well as furthering co-production and participation with renters, rental organisations and municipalities is part of *Eigenhaard's* policy (Eigenhaard, n.d.).

Rochdale

As a housing corporation, *Rochdale's* main aim is to provide sufficient housing for people of all ages, status and origin that are struggling with limited means. Through offering affordable dwellings of good quality they are able to provide 80.000 people with a home in Amsterdam, Zaanstad, Purmerend, Diemen and Landsmeer.

By maintaining the quality and the amount of social housing stock in certain areas, providing additional housing for people with higher incomes, enabling current tenants to climb the housing ladder better and supporting resident initiatives *Rochdale* want to demonstrate their responsibility, thoughtfulness and responsiveness to people' needs (Rochdale, n.d.).

Stadgenoot

Their aim is to provide affordable homes in all parts of Amsterdam. The differentiation of the housing market, meaning less rented accommodations and more home ownership, has led them and other housing associations to ensure that at least 162.000 affordable social rented properties are being retained. *Stadgenoot* themselves offer 26.500 social dwellings in their portfolio and follow the concept of increasing this number. They believe that as the city as a whole is growing, also the social housing

sector must grow. This will ensure the accommodation of vulnerable groups and a multicultural community, which are essential to a city's balance (Stadgenoot, 2016).

Woonzorg Nederland
Woonzorg Nederland have made it their mission to provide housing for vulnerable groups of people and through providing good service, support them in finding safety and security within their accommodations. They especially ensure that their aged or disabled tenants are provided with a dwelling that supports their special needs. Their focus is on people with limited means, that are accommodated through social housing.
On a neighbourhood level *Woonzorg Nederland* ensures innovative, sustainable and flexible provisions of housing, retail, service, spaces of encounter and recreation, to create spaces for carefree living (Woonzorg Nederland, n.d.).

Ymere
Active in Amsterdam, Almere, Haarlem, Haarlemmermeer and Weesp, *Ymer*e provides 70.000 social rental apartments for people with limited means. They are continuously working to improve the quality of their apartments, the range of offered apartment types and to provide new forms of combining working and living. To achieve this they cooperate closely with their tenants and neighbourhood partners such as municipalities and co-makers.
Overall, they are looking to support the metropolitan region Amsterdam in becoming a strong and dynamic region, where their current and future tenants can find affordable homes (Ymere, n.d.).

Cordaan
Cordaan provides living spaces for anyone who requires special care, support or attendance, such as elderly with dementia or physical impairment, children and grown-ups with a mental disability and people with physical problems.
Every year *Cordaan* is able to provide housing for 20.000 people in Amsterdam, Diemen, Huizen and Nieuw-Vennep. With the help of 5.400 employees and 2.000 volunteers, as well as family caregivers, research institutions and social organisations, they have been able to grow to one of the largest care-organisations in Amsterdam and want to take responsibility for establishing long time care in the city (Cordaan, n.d.).

What these brief profiles disclose, is the broad range of aims and target groups each housing corporation has formulated for itself. First and foremost their visions include the provision of affordable housing for target groups having a disadvantage when entering the housing market. These disadvantages can be due to their entry level, income or a physical or mental disability. Their missions then formulate the need to make the access to housing inclusive and on a neighbourhood level balance social, spatial and economic issues to ensure liveable environments for their tenants. How come these straightforward aims are not always achieved, as in the redevelopment of Nieuw West.
This can largely be explained through the policy developments, as well as external circumstances, explained in the following.

7.1.3. Policy Developments

Up until 1990, housing in the Netherlands was largely influenced by the government, as it had been for a long period of time. When early in the 1990s, privatisation of housing associations occured (Boelhouwer, 2007), leading to their financial independence and a change from construction subsidies to rent allowance (van der Veer, 2017).

This early shift of responsibilities and withdrawal of the state with regard to housing matters, is reflected in the first steps to revitalising Nieuw West. In 1993, project initiation took place through the

neighbourhood councils of Osdorp and Geuzenveld-Slotermeer, which established the Bureau Parkstad, without any state involvement (Nio et al., 2016).

Towards the end of the 1990s, the dutch government proclaimed that new housing production should concentrate on the market sector. Determining a broad area of activities for housing associations to be in involved in . As a result, the associations have built only a small amount of social rented housing during the past 5 years, concentrating more on new builds for the market rented sector. Overall, this has created a gap between the need for social rented housing and the supply housing associations could offer (van der Veer, 2004; Boelhouwer, 2007).

In 1995, the situation for the project Nieuw West changed. While the Bureau Parkstadt remained as an important organisational body, the municipality of Amsterdam took on a leading role. From that moment onward, the redevelopment project was closely coordinated with multiple stakeholders, such as the Amsterdam municipal authorities and the neighbourhood councils, Bureau Parkstad, the housing corporations as well as the involvement of other parties such as residents' organisations, welfare organisations, and health insurers, resulting in the formulation of a master plan '*Richting parkstad 2015*' in 2001. (Nio et al., 2016).
Throughout the next phases of the project the collaboration continues. The housing associations focus on their task of providing housing and public facilities, while the Amsterdam municipality realizes public spaces. However in 2005, the realization occurs that the collective development budget of 5 million Euros is coming to an end. Among other things, this leads to the decision to focus more on the improvement of the situation of the current residents, leading to the municipality's move to in 2007, pressure the housing corporations into taking on all the risk. Meaning the associations became solely responsible for the financial risk of demolition and new build, building social housing (90.000m2), and the redevelopment of public space. In return they are left a free hand to act, while the local neighbourhood district is left with determining the area borders (Nio et al, 2016).

After the crisis in 2008, the dutch government established, that 90% of the vacant social housing must be let to households with an income below 35.000 euro/year (Aedes, 2013). At the same time, the Landlord Levy was introduced. A tax to be paid by the associations letting more than 10 dwellings below in the regulated sector at or below the threshold (van der Veer, 2017). This was to regulate the size and scope of the Dutch social housing sector.

In the case of Nieuw West it has the effect of slowing the development down. Which was eventually increased when the investment capacity of housing corporations was reduced, meaning with the *New Housing Act 2015* (nieuwe woningwet) housing associations were not admitted to developing freehold dwellings outside the rent regulated sector anymore. Overall, this policy put into play the idea of moving back to the core activities (van der Veer, 2017).

These developments have resulted in housing associations taking on the role of a more passive player, reacting to the government and external influences. However, as the main housing developer in the case of Nieuw West, housing corporations should have had a more free hand to act according to their visions and implement their missions. Hence, the practical implications to ensure the framework's successful performance are examined. The framework could form the basis of an empowering instrument for housing associations to achieve their social, spatial and economic aims as introduced above, despite government regulations and external influences. The power of holding information on intrinsic local neighbourhood structures could be the key to them occupying a more prominent and dynamic role in planning and decision making phases of urban development and building construction projects.

7.2. The Framework's Managerial Added Value

The process of examining the integrated framework's managerial added value took place through email contact established with the nine housing associations with ownership in Nieuw West.

Each organisation was contacted separately either via email or through the communication platform provided on their website. Included in the email message was the reason for the contact, a brief description of the research and the request to review and complete an attached survey of five minutes.

The survey was setup as a combination of simple filling in of a table and two open ended questions. The table suggested indicators for information collection and the subjects were asked to indicate whether these suggested indicators were of added value to them and their organisations and whether they were already able to collect information on them. The table also left space for additional indicators to be added, which the subjects' organisation perhaps already considered. The table was followed by two open ended questions. The first, asking the subject to depict any other indicator he/she would deem relevant for the organisation to collect. The second, requesting the subject to state his/her thoughts on a big data approach for such information collection and the role of big data analytics as an empowering tool.

Out of the nine housing associations, four responses have been received, of which only two were positive. A brief overview of the contact and responses is given in the table below (table 21) (For further information see Appendix B - Contact and Response Table).

	Response	Positive or Negative	Comment
de Alliantie	No	-	-
de Key	No	-	-
DUWO	No	-	-
Eigen Haard	No	-	-
Rochdale	Yes	P	interested in final research results, as they are building a data set of their own
Stadgenoot	Yes	N	cannot participate
Woonzorg Nederland	No		-
Ymere	Yes	P	interested in final research results, as they are interested in everything that can help them understand their tenants and people of Amsterdam better
Cordaan	Yes	N	cannot participate

Table 21 - Housing Associations' contact and response rate

Finally, the survey responses of Rochdale and Ymere are depicted in a table below, in which the suggested indicators are ranked according to their added value and their collectability (table 22). The ranking was formulated in correspondence with the subjects' positive response to the added value of the suggested indicator and whether the organisation themselves is able to gather information for this indicator. The responses from the two open ended questions are discussed thereafter.

Ranking				
	Highly valued information	Valued Information	Trivial Information	Collectable Information
Suggested Indicator	Median Income			Yes, once at rental start
	Tenant Age			Yes
	Household Size			Yes, once at rental start
	Employment Status			No
	Wish of tenant to move/stay			No
	Satisfaction with the area			No
		Education Level		Yes, on higher scale
		Education Status		Yes, once at rental start
		Marital Status		Yes, once at rental start
		Ethnicity		No
		Personality Types		No, data is bought
			Tenant Gender	-
			Travel time to work, educational facilities	-

Table 22 - Indicator Ranking according to Housing Associations' response

From the 12 suggested indicators only six are deemed highly relevant: *Median Income, Tenant Age, Household SIze, Employment Status, Wish of tenant to move/stay* and *Satisfaction with the area*. The Four indicators *Education Level, Education Status, Marital Status* and *Ethnicity* are only considered second-rate. While the indicators *Tenant gender* and *Travel time to work/educational facilities* are trivial. Only one indicator was added to the suggested one's, *Personality Types*, which is a way of categorizing tenants based on certain characteristics.

From the response to the first open ended question, which asked for the indicators that should be added to the collection to provide a holistic neighbourhood picture, it is determined that overall, any additional information that can help housing associations better understand tenants and the people of Amsterdam, to better accommodate them, is relevant. The wish to classify citizens and comprehend their specific housing needs is central. However, the challenge of respecting people's privacy is a drawback. More Information on tenants' wishes to move or stay and why, what they like about their accommodation and the area or dislike are the central concerns from housing associations' standpoints.

With regard to the second open ended question, concerning the role big data analytics could play in providing such information, it can be derived that housing organisations are eager to explore this opportunity. Generally, they see great potential in big data standards in managing and enhancing decision-making, not only with regard to urban development and building construction projects but on all levels of the organization. While on the one hand, housing organisations immediately imagine the added value of such an approach in practice, e.g. by determining a registered low-income tenant's actual income, through reviewing his social media activity on what kind of car he drives, the concern of crossing the border of their tenants' privacy, remains.

7.2.1. Results of the Survey

Even through only two out of the four received responses proclaimed interest, it is deduced, that the possible opportunities of big data analytics are eagerly received by housing associations, as the two positive reactions were right on enthusiastic. However, it is not really transparent for organisations what exactly current big data standards could improve. This lack of knowledge on big data opportunities is reflected in the little attention that was given to adding more indicators than the suggested ones to the information collection. While more information on the tenants' needs and satisfaction with the area are strongly wished for, the possibilities big data standards bear to support

these wishes are not clear. At the same time, the privacy concerns going hand in and with new big data developments, are high on the housing corporations' radars.

An empowering instrument would without a doubt be strongly embraced by housing associations in the future, as the advantage of such an approach on a practical scale is transparent to them. However, the organizations think of a possible application within the boundaries of managing their core business, while a mature big data instrument could empower them way earlier in the process. They are unaware of the greater added value such an instrument could potentially hold for their organisation, but more importantly for their position during planning and decision making processes and when formulating development objectives with other stakeholders. Therefore, they are so far also oblivious of the all the possibilities the integrated framework poses.

This means, that the managerial added value of the integrated framework as a first step towards a future big data instrument is potentially greater than expected by practitioners, therefore confirming the genuineness of this blueprint.

7.3. Retrieving Practical Recommendations

Retrieving practical recommendations for a future implementation of the integrated framework was achieved through follow up idea exchange and discussion sessions with the two positively responding housing associations. Prior to the meetings the contact persons were provided with an executive summary of the research, in order to provide them with a solid basis of knowledge and understanding of the study based on which the discussion could take place. they were also informed that their input would be specifically required to assess the practical implications of the suggested integrated framework from their perspective as stakeholders in a redevelopment project. The 1-2 hour long meetings were held separately in the each of the organization's headquarters in Amsterdam. To kick-off the meetings a brief presentation was held, stating the relevance of the research, problem statement, main research question, findings and preliminary conclusions. In the following paragraphs the topics, concerns and advantages of the in the framework employed big data standards are summarised.

Individual's Privacy

First and foremost stands the concern of housing associations to protect individual privacy. How can this be ensured when digging through all this data? Naturally, there are strict regulations in place by the platforms providing the data. But in the wrong hands, such detailed information could still be damaging.

Moreover, a new form of discrimination could arise. "Tenant Cherry Picking" based on the information retrieved through big data standards could become a biased form of choosing tenants and promote a new pattern of segregation.

The overarching question that should be asked is "Would I want my data to be openly accessible like that?"

Effects on Company Image

Connected to the above topic is the concern for the company image. What would adversaries of big data analytics say about an organization using big data for their benefit? Could being ''pro big data'' harm the organization's image? Most importantly, what would clients thinks? And is it ethical to use big data standards to retrieve more detailed information? These questions lead back to the overarching statement above "How would I feel knowing my landlord collects data about me used for planning and decision making?"

Effects on Company Competitiveness

Due to these doubts, should an organization renounce from using big data standards for their business entirely? How would this affect the company's competitiveness? Saying no to big data analytics could diminish an organization's position on the market. If the competition is making use of a

big data approach, why should an organisation be self-defeating by swearing off of such an advantage altogether?

Core Value Support

Moreover, the proposed approach supports the core aim of housing associations to accommodate weaker income groups. Being able to establish neighborhood profiles with real-time information and on high resolution, ensures that individual neighborhood needs can be detected an taken into account in the planning and decision making process. This is especially important for less fortunate individuals, as it can be ensured that their support networks, on which they rely so rigorously, are being kept alive. By knowing how a place is used and by whom, a redevelopment plan can build on that existing structure, rather than uprooting establishments. Therefore, the proposed integrated framework is for sure of value, even if individual data is collected.

Implementation of Big Data Approach

Moving toward the actual implementation of such an approach within planning and decision making processes. Apart from the concerns mentioned above, according to the associations, this would take very detailed policy planning from the municipality's side and also from them, to ensure responsible data handling. This is a huge task to be tackled before the process of mining data can actually begin and is complicated by questions of ethics, privacy, information sharing, regulation, procedure and law.

Legal Aspects

The lawfulness of using information retrieved from big data sources is a matter impacting the entire possibility of implementing such an approach. Oftentimes the in the framework proposed data is only released on the condition that it be used exclusively for research. Technically the organizations would also be using it for research, but then adopting the results and basing practical decisions on it. How far is this acceptable from a legal point of view?

As can be seen from the above summary, overall these sessions have lead to more questions. This is progress, as through these meetings prominent concerns as well as the most advantageous aspects have been highlighted.

On the bottom line what housing associations are looking for at this moment, is a way to determine what a "good neighborhood" is. As this is a tricky question that can differ from context to context, a blueprint, as the one proposed in form of the integrated framework, is a welcome starting point for determining if a neighborhood is good, based on indicators such as rental housing supply, key locations and vitality and diversity.

On the other hand, already very concrete ideas of what is wanted from big data analytics exist as well. Scepticism has been voiced with regard to what information social media can truly provide, especially as its use is restricted to a certain population group. Moreover, doubts regarding the applicability of the big data standards with regard to specific causes, such as determining the reason for building decay, have been mentioned. Naturally, the introduced framework is not universally applicable, as it merely depicts a first overview of the possibilities and limitations. Overall, the mindset during these meetings reflected the eagerness to explore all the possibilities big data standards offer, further. With the aim of knowing what one would be deciding for or against and what is even possible before limiting it. In that sense, the integrated framework is useful in providing an overview of the possibilities, but is limited in the direct applicability. It is a first step in the right direction of gathering advantages and concerns to know what one is discussing, but a lot of research is still needed to be able to actually employ such a framework or even develop an instrument that could support planning and decision making processes of urban development and building construction projects.

8. Conclusions

8.1. Responses to Research Queries

This research has been motivated by the observation that current urban redevelopment and building construction projects at times result in unequal or exclusive outcomes and by the curiosity to explore the possibilities of big data standards within this context.

While contemporary state-led interventions, such as revitalization projects, initially aim at dissolving issues of inequality and exclusiveness in the city, through creating an urban social mix and ensuring economic growth, they tend to have negative consequences. Through a literature study it was identified, that the observed negative consequences are able to arise due to a lack of information during planning and decision making processes of such projects. It appears, that during contemporary urban redevelopment projects primarily information regarding *people* and *property* are considered, while data on complex city dynamics, the so-called *area indicators*, are not taken into account.

On the basis of this understanding this thesis pursued to answer the following research question:

How can big data standards be utilized to manage city dynamics and enhance planning and decision making processes of urban development and construction projects in the built environment to mitigate the negative outcomes of such state-led interventions?

Several sub-research questions have been formulated, which have guided the research in answering this main research question. The answers to these sub-research questions have been provided in the individual chapters of this report and are also summarised in the paragraphs below, before answering the main research question.

When reading the following conclusions, it should be considered, that while this research provided valuable insight into the broad range of possibilities and limitations of big data standards in the context of urban development management, the results of this research have not been proven in such a way that they could be generally transferred beyond the investigated context of this study. As the topic of big data overall is so far subject to limited research studies, this thesis understands itself as broadening the knowledge base upon which future research can build, validate and expand.

Sub-research questions
From chapter two onward the questions of *How inequality prevails in cities today?* and *How cities like Amsterdam cope with inequality?* were researched. Inequality in cities today is perceptible in the form of socio-economic inequality, referring to the gap between rich and poor and socio-spatial inequality, meaning the spatial segregation of population groups based on income, education and employment status. Also, the city of Amsterdam is facing such inequality. While overall the issue of socio-spatial segregation is not as immediate as in large North American cities, Amsterdam still faces the issue of so-called poverty concentration areas, which are deprived spaces, within thriving neighborhoods, where disadvantaged groups of the population cluster. What many cities turn to, in order to challenge these inequalities is urban revitalization. It is a form of cleansing the city through physical renewal, aimed at dissolving poverty concentrations through urban social mix and generating economically strong neighborhoods. However, these redevelopment projects have also negative effects. *What are the predominant negative consequences of contemporary urban development and building construction processes?* The main ones being the fact, that such projects oftentimes turn out to be exclusive developments displacing resident groups and thus, relocating the issue instead of solving it. This happens, as contemporary urban development and building construction projects neglect to take into account vital information on city dynamics, when planning and setting development objectives.

Chapter three introduces big data as a new approach toward enhancing the negative consequences of urban development and construction projects in the built environment. *What are the possibilities of big data with regard to such challenges in the urban context?* Big data offers the opportunity of retrieving real-time information on a high resolution. For the urban context this means, revealing how the city is used and by whom, the social connection between people of a place, what certain neighborhood characteristics are or which spaces are attractive compared to others, resulting in detailed information on a neighborhood's social cohesion level and together with the information already taken into account in contemporary developments, providing a holistic neighborhood profile. Naturally, big data analytics also bear drawbacks. *What are the drawbacks of big data analytics?* The main limitations of big data are the issue of depicting correlation but not causality, failing to determine small-number phenomena and creating vicious cycles through the re-use of false data material. Also the challenge of cyber security and privacy concerns are challenges to be taken into account.

Yet the opportunities of big data analytics overrule the limitations, leading to chapter four and the two questions of *How big data can be utilized for the purpose of managing city dynamics?* and *How big data can be utilized for the purpose of enhancing planning and decision making of urban development and building construction processes?* Through using openly accessible and pre-processed data from the platform insideairbnb.com, an independent website offering detailed information on Airbnb in over 30 cities worldwide, information on *Growth of Income, Rental Housing Supply, Increase in Rental Prices, Neighborhood Characteristics, Vitality and Diversity* and *Resident Displacement* in a neighborhood can be provided. From social media data , such as Twitter, Foursquare and FLickr, data on *Key Locations, Attractiveness, Social Cohesion* and the *Usage of the Urban Landscape* can be reviewed. All this data provides vital information for managing ongoings in the city and is specifically useful for planning and decision making process of urban development and building construction projects, in order to tailor development objectives to a neighborhood and its residents' needs.

To achieve this the question *How big data standards capable of managing and enhancing decision making process in the context of urban development and building construction projects look like?* is investigated. Following an illustration based on a case study analysis, an integrated framework is formulated, which combines the examined big data standards into one five-step approach, consisting of indicator determination (step 1), data source identification (step 2), data collection (step 3), data analysis (step 4) and data output (step 5), adding traditional methods of information collection to the blueprint.

Main research query
This integrated framework is the answer to the main research query *How can big data standards be utilized to manage city dynamics and enhance planning and decision making processes of urban development and construction projects in the built environment to mitigate the negative outcomes of such state-led interventions?* as it shows how current big data standards can be combined and incorporated within an integrated model capable of informing planners and decision makers with regard to city dynamics in the context of urban development and building construction projects and ensuring the possible negative consequences of such state-led interventions are mitigated.
However, due to the mechanism of the data sources it is vital to highlight, that current big data standards are not capable of doing this alone. The fact that social groups are excluded and some details are correlated in ways, where the theoretical results do not uphold in practice are prominent issues. The integrated framework proposes an approach, that is almost ahead of time - when *Generation C* has their own children and the beforehand mentioned global trends are progressing, more and more parts of the population will be connected and thus represented in the proposed data sets.

On top of the integrated frameworks information power, it is also investigated whether *Big data standards are able to empower practitioners within planning and decision making processes?* and *What are the practical requirements to implement such standards in practice?* Clearly, the information that can potentially be provided through the suggested framework, would enable passive stakeholders to seize a more prominent role during the formulation process of development objectives, and base a discussion on detailed knowledge of a neighborhood and its residents. However, to implement this or a similar instrument in practice, detailed policy planings are required and many more questions and concerns need to be faced.

8.2. Relevance of this Work for Theory and Practice

The aim of this research has been to investigate, whether big data standards hold opportunities to enhance the management of urban development and building construction projects and thus mitigate the negative consequences. This was aimed at achieving through a wide exploration of literature on the topics of state-led revitalization and big data standards, in order to create an extensive overview over the possibilities of big data within this context. The result of this literature exploration was the integrated framework, which depicts the information vital to create holistic neighborhood profiles, a combination of big data and traditional indicators.

As such, this research is of great informative value for both theory and practice of the arising digital age. As mentioned, this report holds the knowledge base on which future research in this field can build upon. For practice this means, parties interested in a big data approach can review the extensive possibilities, as well as limitations, when considering to implement big data as an instrument in the profession in the future.

8.3. Limitations and Further Recommendations

Naturally, this research bears some limitations. First of all, the blueprint has only been illustrated on one case study. It would therefore, be highly recommended to illustrate the framework performance through further case studies, which would help substantiate the results of this research further.

Moreover, this study was conducted by only one researcher, limiting the perspective, but also the expertise with which this topic was approached. Due to the complexity of the topic of big data, it is highly recommended that the outcomes be verified and the study continued by researchers with different perspectives.

Not only the research itself but also the components subject to this study bear limitations. Overall, it was discovered, that information retrieved through big data analytics alone is unable to provide conclusive result. Most of the time, this data is correlated, cross-referenced and re-checked against other data and information from traditional data sources. This restraint also has to do with the limitations of each individual data source. Only a fraction of the whole volume of data is made available and is subject to random pre-selection. Furthermore, the platforms from which the data is retrieved are used by a specific age and social group of the population, which narrows the information collection pool. This limitation is tried to diminish by integrating information from traditional sources, such as census and registry data.

Apart from that, the research outcome is dependant on a data-rich environment meaning, that in areas where data is scare the framework's performance is limited. This has been proven with the illustration in the context of the case study analysis. It is therefore recommended, to proceed with further illustrations to substantiate the results.

Moreover, establishing the managerial added value of the integrated framework was subject to time constraints. The interest in a blueprint for information collection, such as the integrated framework, was confirmed, by only two out of four responses from nine contacted organizations. While there, was no sufficient time to follow up with all housing associations, it is also questionable, what exactly planners and decision makers would need to employ big data standards. From the discussion

sessions more questions than answers have been raised, which once again reflects the immaturity of this topic. Also, again the limited number of participants was a drawback. For optimal results, it is suggested to hold such sessions with all concerned parties, first separately and then a follow up group sessions, from which a joint plan of action could be brought forth to establish the next steps toward implementing big data standards as information collection methods.

Finally, the aim of exploring big data possibilities in the context of urban development and construction projects in the built environment has been successfully achieved. The opportunities and drawbacks of big data analytics have been thoroughly examined. Based on the conclusions of this examination, it was possible to formulate an integrated framework, which is of great informative value and forms the basis for a future instrument capable of enhancing the management of contemporary urban development and construction projects. On top of that the frameworks potential managerial added value was confirmed through practitioners. Overall, this work is considered to be of great added value in a more connected future.

9. Reflection

How the research sits within the master program
This master thesis on exploring the possibilities of big data standards to enhance the management of planning and decision making processes of urban development and construction projects in the built environment is part of the graduation lab *BOLD Cities*. Belonging to the master track Management in the Built Environment, this graduation lab concerns the exploration of new means for urban development management and construction processes, to ensure high quality developments, which are also financially rewarding. Within the master program Architecture, Urbanism and Building Technology, this research is centered around managerial dimensions and processes of the built environment and the construction industry.

The choice of research method
Exploring the application of big data as a research method was predetermined by the graduation lab, as the aim of the lab coordinators was to give students the opportunity of investigating this quite young and unexplored topic. Specifically, the use of big data analysis to retrieve real-time and high resolution information on city dynamics in order to manage and enhance planning and decision making, is a so far uncharted field. Therefore, my particular research offered the prospect of exploring the opportunities of big data analytics in the context of urban development and construction projects in the built environment, with regard to the planning and decision making phase.

The literature reviewed for this research concentrated on two areas. First, it was focused on the societal challenge of inequality in cities and how the means to mitigate these inequalities lead to negative consequences themselves. Moreover, the reasons for these negative consequences were determined, from which a problem statement was formulated. Secondly, literature on the scientific challenge of laying out the possibilities and limitations of big data analytics with regard to solving this specific matter was reviewed and relevant data sources were determined.
The aim of this research was then to explore, whether a big data could be utilized to enhance the management of planning and decision making process of urban development and building construction projects to ensure more inclusive development outcomes, which are socially mixed and promise economic neighborhood growth.

Structure and limitations of the approach
During the research process it was discovered, that information retrieved from big data analytics alone is not able to produce conclusive results. Oftentimes, information is cross checked against traditionally acquired information or correlated with other data. This lead to the research outcome of developing an integrated framework, which incorporates big data standards of information collection with traditional means. This integrated framework is an attempt to close the information gap planners and decision makers of contemporary urban development and construction projects are facing when formulating development objectives, by ensuring a holistic neighborhood profile is established.
When illustrating the framework performance through a case study in Amsterdam Nieuw West, another limitation of the blueprint is discovered. The big data standards of the framework are only able to produce unambiguous data output, when applied in a data-rich environment. Unfortunately, the data volume in Nieuw West was, to a certain degree, insufficient, resulting in parts of the retrieved data being trivial. In order to ensure the integrated framework performs successfully, meaning it provides conclusive and holistic data on urban dynamics, the outcomes need to be substantiated through further illustrations.
The integrated framework's added value for the management practice is included into the research as well. By consulting practitioners with regard to the abilities of the blueprint for enhancing the management of their core business and empowering them outside of that, e.g. in negotiations with other stakeholders, the potential of the framework in the professional field was explored. Initially, the

nine housing associations in Nieuw West were selected to participate in a survey. Their expertise is considered of value to my research as they were key stakeholders in the case study. The aim of the survey was to establish the status quo of information collection and potential of big data, from the organizations' viewpoints.

In the available time span, four replies were collected, of which two were positive, filling in the survey and adding their thoughts on big data analytics as an empowering tool. Due to the time limitation, also the planned follow up idea exchange and discussion sessions could only be scheduled with the two positive responding parties. For more conclusive results, it would be optimal to have responses and discussion sessions with all nine housing associations. As mentioned, time was not on my side here.

Overall, the integrated framework depicts the broadened knowledge base, with regard to big data in the urban development context

With regard to encountered ethical issues, privacy was naturally a big concern. Yet, being the main concern when it comes to big data analytics, it seems, that it is being thoroughly considered. Especially, platforms, that are making data available for research, are conscious about protecting the privacy of others, e.g. through anonymizing data, providing random samples of data, releasing only 1% of data etc. Once the data is in the hands of researchers, it is their responsibility to handle sensitive data with care and comply with the terms and conditions of use.

Mentor feedback

Due to the fact that the *Bold Cities* graduation lab is a completely new theme at the MBE department and it was clear from the start, that research would be very explorative, the mentors left a quite free hand. At times they took the opportunity to reign all ideas back in and help frame the research. In my case, by pointing me toward deciding what the focus points are and what the overall question of the research is, as opposed to what the application area is, and such matters. To name one concrete example, at the P2 the number of case studies I proposed, was discussed. It was recommended by the mentors to focus on one case as an application area to my tool. I embraced this recommendation at it made sense to focus my demonstration in detail on one case, as the added value of demonstrating the tool's applicability on two cases was not of significance and might even hinder the aim of going in depth.

Overall, this 'complete freedom' was new and at times it was difficult to determine what to do next and one wished for more supervision. Despite this, I believe that it was necessary, in order for this research to be truly explorative, that no definite boundaries were set from the beginning.

On the other hand, at the P2 and P3 the feedback was so rich and a lot of recommendations were made, that it was necessary to choose and focus on a few that were relevant to the aim and scope of my research.

What I learned

From the way in which this graduation lab was held, I learned how to trust more in the work I am doing and the decisions I am making. But also, how to stand for the choices I make, knowing when and how to defend them and when to be open to adapt them. Specifically, the topic and work I was doing enabled me to grasp the abstract concept of big data and translate it into something beneficial for my future profession. Additionally, I gained more insight into the societal matters of urban development management, which I value highly.

10. References

Aedes - vereniging van woningcorporaties (2013) *Dutch social housing in nutshell*, Aedes Brussels, Housing Europe Center, EU public affairs

ArchEX (n.d.) *The Western Garden Cities*, retrieved on 6th December, from http://archex.info/english/netherlands/amsterdam_western_garden_cities.html

Arribas-Bel, D. and Sanz-Gracia, F. (2014) *The validity of the monocentric city model in a polycentric age: US metropolitan areas in 1990, 2000 and 2010*, Urban Geography, 35:7, 980-997, DOI: 10.1080/02723638.2014.940693

Atkins, R. (2002) *Does Gentrification help or harm Urban Neighbourhoods? An Assessment of the Evidence-Base in the context of the New Urban Agenda*, ESRC Center for Neighbourhood Research, Paper no. 5

Bendler, J., Brandt, T., Neumann, D. (2015) *Does Social Media Reflect Metropolitan Attractiveness? Behavioural Information from Twitter Activity in Urban Area*s, Business Analytics Congress, Fort Worth 2015, DOI: 10.13140/RG.2.1.3272.13

Byrne, J.P. (2003) *Two Cheers for Gentrification*, Georgetown University Law Center, Howard Law Journal, Vol. 46 No. 3

Boelhouwer, P.J. (2007) *The future of Dutch housing associations*, Journal of Housing and the Built Environment, December 2007, Volume 22, Issue 4, pp 383–391, 22: 383. https://doi.org/10.1007/s10901-007-9091-8

Cameron, S. (2003) *Gentrification, housing redifferentiation and urban regeneration: 'going for growth' in Newcastle upon Tyne*, Urban Studies, 40, pp. 2367–2382.

CBS Statistics Netherlands (n.d., A) *Organisation*, retrieved on 9th April 2018, from https://www.cbs.nl/en-gb/about-us/organisation

CBS Statistics Netherlands (n.d., B.) *Dutch Census Saves Time and Money*, retrieved on 9th April 2018, from https://www.cbs.nl/en-gb/news/2014/47/dutch-census-saves-time-and-money

Coldwell, W. (2014) Airbnb's legal troubles: what are the issues? in The Guardian International Edition, retrieved on 21st March, 2018, from https://www.theguardian.com/travel/2014/jul/08/airbnb-legal-troubles-what-are-the-issues

Cordaan (n.d.) *Over Cordaan*, retrieved on 1st April 2018, from https://www.cordaan.nl/over-cordaan

Cranshaw, J., Schwartz, R., Hong, J.I., Sadeh, N. (2012) *The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City,* School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, Association for the Advancement of Artificial Intelligence (www.aaai.org).

Cucca, R. (2012) The Unexpected Consequences of Sustainability. Green Cities Between Innovation and Ecogentrification, Sociologica, 2/2012, Societa editrice il Mulino, Bologna, DOI: 10.2383/38269

de Alliantie (n.d.) *Over de Alliantie*, retrieved on 1st April 2018, from

https://www.de-alliantie.nl/over-de-alliantie/wie-zijn-we/

de Key (n.d.) *Visie Amsterdam*, retrieved on 1ˢᵗ April 2018, from https://www.dekey.nl/#!PageID=4107&content=&contentid1=&contentid2=

Dienst Ruimtelijke Ordening Amsterdam (2012) C-Change Regional Report 2009-2012, retrieved on 9th December 2017

Divercities (2017) *Rotterdam*, retrieved on 24th November 2017, from https://www.urbandivercities.eu/rotterdam/

Dudas, G., Vida, G., Kovalcsik; T., Boros, L. (2016) *A socio-economic analysis of Airbnb in New York City*, Regional Statistics, DOI: 10.15196/RS07108

DUWO (n.d.) *The history of DUWO*, retrieved on 1ˢᵗ April 2018, from https://www.duwo.nl/en/about-duwo/the-organisation/the-history-of-duwo/

Eigenhaard (n.d.) *Dit is Elgenhaard*, retrieved on 1ˢᵗ April 2018, from https://www.eigenhaard.nl/eigen-haard/wie-zijn-we/dit-is-eigen-haard

EUKN - European Urban Knowledge Network (2010) Renovation operation in Amsterdam Nieuw West, retrieved on 19th March 2018, from http://www.eukn.eu/e-library/project/bericht/eventDetail/renovation-operation-in-amsterdam-nieuw-west/

Fenton, A. (2016) *Gentrification in London: A progress report 2001-2013*, London School of Economics, UK

Florida, R. (2017) *The New Urban Crisis: how our cities are increasing inequality, deepening segregation, and failing the middle class - and what we can do about it*, Basic Books, Hachette Book Group, Inc, New York 2017

Foursquare (2017) *Platform Policy*, retrieved on 26th March 2018, from https://de.foursquare.com/legal/api/platformpolicy

Galster, G., Peacock, S. (1985) *Urban Gentrification: Evaluating Alternative Indicators*, College of Wooster, U.S.A.

Gant, A.C. (2016) *Holiday Rentals: The New Gentrification Battlefront*, Sociological Research Online, 21 (3), 10, DOI: 10.5153/sro.4071

Gemeente Amsterdam (n.d.) Social housing and private sector rentals, retrieved on 9th April 2018, from https://www.amsterdam.nl/en/housing/rental-prices/

Glaeser, E L., Kominers, S.D., Luca, M., Naik, N. (2015) *Big Data and Big Cities: The Promises and Limitations of Improved Measures for Urban Life*, Faculty Research Working Paper Series, Harvard Kennedy School, John F. kennedy School of Government

Gurran, N., Phibbs, P. (2017) *When Tourists Move In: How Should Urban Planners Respond to Airbnb?*, Journal of the American Planning Association, 83:1, 80-92, DOI: 10.1080/01944363.2016.1249011

Hall, P. (1997) *Modelling the post-industrial City*, Futures Vol 29, No.4/5, p.311-322

Hall, P. and Pain, K. (2006) *The Polycentric Metropolis Learning from Mega-City Regions in Europe*, Earthscan, London/Sterling, V/A

Hristova, D., Williams, M.J., Musolesi, M. Panzarasa, P., Mascolo, C. (2014) *Measuring Urban Social Diversity Using Interconnected Geo-Social Networks*, International World Wide Web Conference Committee, ACM 978-1-4503-4143-1/16/04. , http://dx.doi.org/10.1145/2872427.2883065

Iamsterdam (n.d., A) *Geuzenveld-Slotermeer*, retrieved on 28th March 2018, from https://www.iamsterdam.com/en/living/about-living-in-amsterdam/neighbourhoods/geuzenveld-slotermeer

Iamsterdam (n.d., B) *Osdorp*, retrieved on 28th March 2018, from https://www.iamsterdam.com/en/living/about-living-in-amsterdam/neighbourhoods/osdorp

Iamsterdam (n.d., C) *Slotervaart-Overtoomse Veld*, retrieved on 28th March 2018, from https://www.iamsterdam.com/en/living/about-living-in-amsterdam/neighbourhoods/slotervaart-overtoomse-veld

Inside Airbnb (n.d.) About Inside Airbnb, retrieved on 2nd March 2018, from http://insideairbnb.com/about.html

ISHF - International Social Housing Festival (n.d.) The renewal of the Western Garden Cities: lectures + guided bike tour, Cultural heritage Agency of the Netherlands, retrieved on 22nd March 2018, from https://www.socialhousingfestival.com/events/renewal-western-garden-cities-social-housing-heritage-meeting/

Kelley, P.G., Sleeper, M., Cranshaw, J., (2013) *Conducting Research on Twitter: A Call for Guidelines and Metrics*, CSCW Measuring Networked Social Privacy Workshop 2013

Kitchin, R. (2013) *The real-time city? Big data and smart urbanism*, GeoJournal (2014) 79:1–14, Springer, DOI 10.1007/s10708-013-9516-8

Kühn, M. (2015) *Peripheralization: Theoretical Concepts Explaining Socio-Spatial Inequalities*, European Planning Studies, 23:2, 367-378, DOI:10.1080/09654313.2013.862518

Lees, L. (2008) *Gentrification and Social Mixing: Towards an Inclusive Urban Renaissance?*, Urban Studies 45(12) p. 2449–2470, DOI: 10.1177/0042098008097099

Le Roux, Y. (2012) *Privacy and Data Analytics*, CA Technology Exchange, Insights from CA Technologies, Issue 2, Vol.3, p.17-23

MO, (2013) *Retracing the Garden City*, Het levende archief van Merijn Oudenampsen, socioloog en publicist, blog post, retrieved on 9th December 2017, from https://merijnoudenampsen.org/2013/04/03/retracing-the-garden-city/

Moseley, N., Ovesdotter Alm, C., Rege, M. (2015) *On Utilizing Nonstandard Abbreviations and Lexicon to Infer Demographic Attributes of Twitter Users*, T. Bouabana-Tebibel and S.H. Rubin (eds.), Formalisms for Reuse and Systems Integration, Advances in Intelligent Systems and Computing 346, DOI: 10.1007/978-3-319-16577-6_11

Musterd, S. (1994) *A rising European underclass? Social polarisation and spatial segregation in European cities*, Built Environment Vol 20, No.3, p.184-191

Musterd, S., Marcińczak, S., van Ham, M., Tammaru, T. (2017) *Socioeconomic segregation in European capital cities. Increasing separation between poor and rich*, Urban Geography, 38:7, 1062-1083, DOI: 10.1080/02723638.2016.1228371

Nio, I., Reijndorp, A., Veldhuis, W., Blom, A., Coumou, H. (2016) *Nieuw West: parkstad of stadswijk - De vernieuwing van de Westelijke Tuinsteden Amsterdam*, transcity valiz, Nederland

Paldino, S., Bojic, I., Sobolevsky, S., Ratti, C., González, M.C. (n.d.) *Urban Magnetism Through The Lens of Geo-tagged Photography*, https://arxiv.org/pdf/1503.05502.

Pararius (2018) Rental Apartments Amsterdam (Stadsdeel Nieuw-West), retrieved on 9th April 2018, from https://www.pararius.com/apartments/amsterdam/city-district-stadsdeel-nieuw-west

Quattrone, G., Proserpio, D., Quercia, D., Capra, L., Musolesi, M. (2016) *Who Benefits from the "Sharing" Economy of Airbnb?,* International World Wide Web Conference Committee (IW3C2), April 2016, Montréal, Québec, Canada

Rabari, C. and Storper, M. (2013) *The digital skin of cities: urban theory and research in the age of the sensored and metered city, ubiquitous computing and big data,* Cambridge Journal of Regions, Economy and Society 2015, 8, 27–42 doi:10.1093/cjres/rsu021

Rochdale (n.d.) *Misie en strategie*, retrieved on 1st April 2018, from https://www.rochdale.nl/missie/

Salvati, L. (2017) *Demographic transition, immigration, gentrification: Unravelling early signs of re-urbanisation in a European city*, Council for Agricultural Research and Economics (CREA), Rome, Italy

Sans, A. A., Dominguez, A.Q.. (2016) *Unravelling Airbnb: Urban Perspectives from Barcelona*, Journal Reinventing the Local in Tourism: Producing, Consuming and Negotiating Place; Channel View Publications: Buffalo, NY, USA; Part 3: Built Environments and 'Glocalized' Spaces; p. 209

Schuiling, D., van der Veer, J. (2004) Governance in Housing in Amsterdam and the Role of Housing Associations, Paper for the International Housing Conference in Hong Kong 2004, organised by the Hong Kong Housing Authority, 2-4 February, 2004

Smith, N. (1996) *The new urban Frontier - Gentrification and the revanchist city*, Routledge, London, UK

Sobolevsky, S., Bojic, I., Belyi A., Sitko, I., Hawelka, B., Murillo Arias, J., Ratti, C. (2015) *Scaling of city attractiveness for foreign visitors through big data of human economical and social media activity*, 2015 IEEE International Congress on Big Data

Stadgenoot (2016) *Over Stadgenoot*, retrieved on 1st April 2018, from https://www.stadgenoot.nl/over-stadgenoot/english/

Stadslente (2016) *De verniewing van Niew-West*, blog post by Gerben Helleman, retrieved on 10th December, 2017, from http://stadslente.blogspot.nl/2016/12/de-vernieuwing-van-nieuw-west.html

Tene, O. and Polonetsky, J. (2013) Big Data for All: Privacy and User Control in the Age of Analytics, 1 Nw. J. Tech. & Intell. Prop. 239

University College London (2012) The INEQ-Cities Atlas, retrieved on 9th April 2018, from https://www.ucl.ac.uk/ineqcities/atlas

Uitermark, J. (2003) *Social mixing and the management of disadvantaged neighbourhoods*, Urban Studies, 40, pp. 531–549

University College London (n.d.) *Amsterdam: Socio-Economic Indicators*, retrieved on 12th December 2017, from https://www.ucl.ac.uk/ineqcities/atlas/cities/amsterdam/amsterdam-sei

van der Bijl, V. (2016) *The effect of Airbnb on house prices in Amsterdam,* Master Thesis, MSc Business Economics: Real Estate Finance & Corporate Finance, University of Amsterdam (UvA)

van Esteren Museum (n.d., A) *Geuzenveld*, retrieved on 28th March 2018, from http://vaneesterenmuseum.nl/en/garden-cities/geuzenveld/

van Esteren Museum (n.d., B) *Slotermeer*, retrieved on 28th March 2018, from http://vaneesterenmuseum.nl/en/garden-cities/slotermeer/

van Esteren Museum (n.d., C) *Osdorp*, retrieved on 28th March 2018, from http://vaneesterenmuseum.nl/en/garden-cities/osdorp/

van Esteren Museum (n.d., D) *Slotervaart*, retrieved on 28th March 2018, from http://vaneesterenmuseum.nl/en/garden-cities/slotervaart/

van Esteren Museum (n.d., E) *Overtoomse-Veld*, retrieved on 28th March 2018, from http://vaneesterenmuseum.nl/en/garden-cities/western-city-gardens/overtoomse-veld/

van Gent, W.P.C., (2011) *Urban Renewal and Social Mixing in Bijlmermeer, Amsterdam*, Presentation Seminar "Urban Renewal Projects in Europe" Institut dÚrbanisme de Paris, retrieved on 9th December 2017, from http://urbanclinic.huji.ac.il/sites/default/files/case%20study%20Biljmammer%20social%20mixing.pdf

van Gent, W.P.C., (2013) *Neoliberalization, Housing Institutions and Variegated Gentrification: How the 'Third Wave' Broke in Amsterdam*, International Journal of Urban and Regional Research, Vol. 37.2, p. 503-522, DOI:10.1111/j.1468-2427.2012.01155.x

van Heelsum, A. (2007) *Case Study on Housing - Amsterdam Netherlands,* European Foundation for the Improvement of Living and Working Conditions, Wyattville Road, Loughlinstown, Dublin 18, Ireland

Wassenberg, F. (2006) *The integrated renewal of Amsterdam's Bijlmermeer high-rise*, Informationen zur Raumentwicklung Heft 3/4.2006, p.191-202

Wheeler, B., Shaw,M., Mitchell, R., Dorling, D. (2005) *LIFE IN BRITAIN - Using millennial Census data to understand poverty, inequality and place*, Technical report, University of Sheffield, UK

Woonzorg Nederland (n.d.) *Waar wij voor staan*, retrieved on 1st April 2018, from https://www.woonzorg.nl/over_ons/wie_wij_zijn/waar_we_voor_staan_

World Population Review (2018) Amsterdam Population 2018, retrieved on 9th April 2018, from

http://worldpopulationreview.com/world-cities/amsterdam-population/

Ygitcanlar, T., Teriman, S. (2013) *Rethinking sustainable urban development: towards an integrated planning and development process*, Int. J. Environ. Sci. Technol. (2015) 12:341–352, DOI 10.1007/s13762-013-0491-x

Ymere (n.d.) *Ons Bedrijf,* retrieved on 1st April 2018, from https://www.ymere.nl/over-ymere-2/ons-bedrijf.html

Yrigoy, I. (2016) *The impact of Airbnb in the urban arena: towards a tourism-led gentrification? The case-study of Palma old quarter (Mallorca, Spain)*, In: Blàzquez, M., Mir-Gual, M., Murray, I. y Pons, G.X. (eds.). Turismo y crisis, turismo colaborativo y ecoturismo. XV Coloquio de Geografía del Turismo, el Ocio y la Recreación de la AGE. Mon. Soc. Hist. Nat. Balears, 23: 281-289. SHNB-UIB-AGE. ISBN 978-84-617-5115-0.

11. Appendix

Airbnb Data Research	Expectations	Assumption	Method	Limitations	Added Value (for my research)	Tasks	Solution/How to	Applicability
1) Dudas et al (2017) How socioeconomic factors of neighbourhoods impact the distribution of Airbnb rentals	Indentifying neighbourhood characteristics	Determine whether a neighbourhood is gentrifying based on amount of Airbnb rentals	1. determine amount of listings in an area from Inside Airbnb 2. determine socio-economic factors of the area through census data 3. determine attractiveness of the are through data from OpenStreetMap/geonbmk de 4. correlate these data sources with pearson cross-correlation matrix	1. Inside Airbnb: see E7 additionally only a snapshot is reviewed at a certain point in time 2. Census data: see E14	1.+2. can be applied to my research 3. depends on access to data for my area 4. depends on my expertise	check whether data from 3 is accessible for Amsterdam Nieuw West and find out if making a pearson cross-correlation model is doable	3. is accessible if I learn how to use QGIS	YES
2) Gurran Phibbs (2017) Impact of Airbnb rentals on neighbourhood concerning noise, congestion, competition for parking, reduction in permanent rental housing supply increase in rental prices	Indentifying neighbourhood characteristics: rental housing supply and increase in rental prices, added source/growth of household income	determine whether a neighbourhood is gentrifying based on it's characteristics, the supply of rental housing and rental prices and growth in household income impact on household income	1. determine neighbourhood characteristics (noise, congestion, parking issues) through examining how tourism and residential activities are regulated, review state and local controls and regulations to certain areas. 2. determine supply of rental housing and rental prices through examining the amount of airbnb rentals (point-in-time information on the number of rentals in the area, the advertised availability, the number of Airbnb listing per host, nightly prices and reviews per month and composition of the listings e.g. whole houses, apartments, rooms, shared rooms) in an area with Inside Airbnb data 3. determine impact on rental housing supply through rental vacancy rates from permanent rental advertisements data 4. determine added income through comparing census data on median monthly mortgages and state government data on median monthly rents with the average monthly incomes from each Airbnb accommodation type	2. Inside Airbnb data: potential for outdated or 'bait and switch' listings, no ability to determine for how long the listings have been available for short-term renting, no availability to determine whether they are conventional B&Bs now using Airbnb platform for marketing and booking 4. Census data: only median amounts of monthly mortgages are reported. These may be lower than new mortgages and loans held by first-time homebuyers.	2.+4. can definitley work in my research. access to Airbnb data is given, census data can be obtained as well. 1.+3. depends on a) access to data regarding land use regulations for tourism and residential activities and b) access to permanent rental advertisement data	check whether data for 1 and 3 is accessible in Amsterdam Nieuw West	1 according to Coldwell, 2014 Amsterdam has not that much legal trouble with Airbnb due to strict regulations of tourism activities: the city of Amsterdam states that from January 2019 the maximum amount for residents to rent out their homes will be reduced to 30 days a year (Gemeente Amsterdam, 1st January 2018) 3. Vacancy can be determined from advertised rentals on funda and compared to the Dutch average, tendency that vacancy is higher in tourist areas	YES
3) Yrigoy (2016) Impact of Airbnb on the rurban area. 3 assumptions: Airbnb enriches low-income population, Airbnb invigorates housing market, Airbnb boosts tourism	Indentifying income/growth of income, impact on housing market, social structure of the area (popularity)	determine whether a neighbourhood is gentrifying based on supply of rental housing, rental prices and growth in income through Airbnb rentals	1. determine supply of rental housing through reviewing Airbnb data specifically the typology of units rented (whole apartments for rent likely that their main purpose is not everyday life but realizing the exchange value of the property) and the ownership/hosts of listings (one host with many listings, likely to be a corporation/rental agency). 2. determine impact on rental prices through reviewing Airbnb rental prices. 3. added income is determined by owner/hosts and rental prices of Airbnb (the more listings from one host, probably not an added income but a main source as renting is run as a business)	Paper does not mention limitations, but for Airbnb data they will be the same as above	All can be applied in my research as access to Airbnb data in my area is given.			YES
4) Quattrone et al. (2016) Who benefits from the sharing economy of Airbnb	Identifying socioeconomic conditions of areas benefiting from Airbnb	Determine whether a neighbourhood is gentrifying based on the enhanced socioeconomic conditions of a neighbourhood benefitting from Airbnb rentals	1. determine Airbnb listings per area from airbnb com (ratio between number of Airbnb rentals in a given area and the km2 of that area 2. determine Airbnb demand per area from airbnb com (using amount of reviews per area as a proxy for demand) 3. determining socio-economic conditions through census data (population density, number of young people, number of educated people, proxies concerning pleasantry of areas, housing information: number of flats and houses present, number to properties sold, number of dwellings overdeliveried, median house price) and through Gini-Simpson-index (ethnicity index: probability that two individuals chosen at random from an area are of different ethnicity) and through the Cultural Index (bohemian index: fraction of people employed in arts, entertainment, recreation) and through the Melting Pot Index (foreigner index: number of people born outside the country divided by the total number of people in the area) and through Index of Multiple Deprivation (small level census: income, employment, health, education, barrier to housing services, crime, living environment) 4. determining attractiveness of an area through transport accessibility from census and through Fourisquare and through Ordnance Survey (captures areas that are covered by attractions normalized by the size of are in km2)	3. census data: only cross-sectional data not longitudinal, makes it difficult to study causal mechanisms.	1. can be applied but using data from insideairbnb. com for listings per area 2. can be applied as is 3. census data can be used for determining socio-economic conditions: Ethnicity index can be derived form census, cultural index as well, Melting Pot also. Multiple Deprivation Index is Leefbarometer in NL, can be applied 4. Foursquare is not applicable in the area, Ordnance Survey is not applicable.	check for 4. attractiveness if Instagram is an option and if there is something like Ordnance Survey in NL	4. attractiveness through POI from www.avewoncities. com, BUT no attractions in Niew West: accessibility is part of attractiveness determined through NS.nl	PARTLY
5) Gant (2016) Holiday rentals: The new gentrification battlefront	Indentifying areas of resident displacement through holiday rentals attracting tourists	Determine whether a neighbourhood is gentrifying/experiencing displacement based on the amount of holiday rentals (Airbnb)	1. determine displacement through examining the amount of Airbnb listing in an area compared to the total housing stock and the type of listing in an area from insideairbnb com	1. Limitations same as above see (E8)	1. Inside Airbnb data is open accessible, this can be done in my research			YES
6) Sans Dominguez (2016)								

Unravelling Airbnb: urban perspectives from Barcelona; Urban impacts of Airbnb on revitalizing neighbourhoods, increasing and diversifying neighbourhoods, providing economic support for families	Identifying urban impacts of Airbnb on neighbourhoods and individuals	Determine the revitalization, diversification of a neighbourhood (=gentrification) and growth of income of families through Airbnb rentals	1. determining **revitalization** through Airbnb rentals by overlapping Airbnb locations with traditional hotel locations to see whether Airbnb are located in areas without hotels/less hotels and by comparing the amount of rentals in an area with the migration rate taken from Census data (more densely clustered Airbnb likely more migration) and through examining listings descriptions to determine the genuity of the experience of that area 2. determining **diversification of a neighbourhood** by examining host origins (through comparing host languages with population groups speaking that language registered in the area) 3. determining **growth of income** through the Airbnb supply of an area and the average income of that area, the characteristics of family households in that area and the number of listings and characteristics of listings offered by the host in that area	Paper does not mention limitations, but for Airbnb data they will be the same as above, same for Census data	All can be applied in my research as access to Airbnb data and Census data in my area is given.	**YES**

SUMMARY						
	Conclusion Assumptions	**Conclusion Limitations**	**Added Value in my research**			
	Based on assumptions it is concluded that Airbnb Data can help determine whether a neighbourhood is gentrifying based on the amount of Airbnb rentals in an area which impact **the supply of rental housing, rental prices, growth in family income and socioeconomic conditions of the neighbourhood.**	**Airbnb Data:** potential for outdated or 'bait and switch' listings, no ability to determine for how long the listings have been available for short-term renting, no availability to determine whether they are conventional B&Bs now using Airbnb platform for marketing and booking ; additionally only a snapshot of data as of a certain date is reviewed **Census data:** only median amounts of monthly mortgages are reported. These may be lower than new mortgages and loans held by first-time homebuyers additionally only cross-sectional data not longitudinal makes it difficult to study causal mechanisms	Use **Airbnb data to** A) determine **displacement/migration rate** in an area (through amount of listings in an area) A1) cross check with demand for Airbnb listings in an area through reviews per area B) determine the the **supply of rental housing and rental prices** (through amount of listings) in area C) determine **added rental income** for households through amount of listings per host and the amount of listings in an area cross checked against the average income of that area D) determine **revitalization** through overlapping Airbnb locations with hotel rentals E) determine **diversification** of an are through examining host origins (compare host languages with census data registered residents speaking that language)			
			General Tasks			
			Go in depth about detailed limitations of Airbnb data from insideairbnb.com			
			Next Step:			
			I am curious to explore whether these assumption are valid and whether an added value is achieved. In other words, how far this data can contribute to mitigating the negative consequences of contemporary gentrification policies.			

A.II. Social Media Data Assumptions

Social Media Data Research	Expectations	Assumption	Method	Limitations	Added Value (for my research)	Tasks	Solution/How to	Applicability
1) Hristova et al. (2014) — Social diversity of urban localities through the social network and mobility patterns of their visitors	Identifying social cohesion and gentrification (places that bring together strangers vs. places that bring together friends)	Determine whether a neighbourhood is gentrifying/its level of social cohesion based on network of Foursquare places and parallel Twitter social network of visitors through check-ins	1. determine the social network (user) of individuals through their direct Twitter network (if A follows B and B follows A there is a social network) 2. determine the place network (venue) through the check-in history of individuals (if A went to X and then to Y and B and C do the same there is a place network) 3. determine the type of venue and its geographical location through Foursquare characteristics 4. determine deprivation of areas through Index of Multiple Deprivation (IMD) 5. Correlating the information on venues reveals bridging places (between strangers) and bonding places (between friends) indicators of brokerage and serendipity 6. Correlation the information on users and deprivation reveals urban social mix and indicators of entropy and homogeneity 7. brokerage, serendipity, entropy and homogeneity together indicate social cohesion levels (tendency the more diverse and deprived the more prone to gentrification)	Paper does not directly discuss limitations; how an area really feels and what its social added value is can only b determined through interviews Foursquare data is not publicly available and therefore derived from Twitter where is publicity accessible PROBLEM??	1. + 2. should be openly accessible but how to collect this info on one area? is Foursquare used in Amsterdam? 3. no issues 4. IMD is leefbarometer in NL, no issues 5. -7. correlating the data is possible if 1.+2. work out	Find out if how Twitter data can be collected for one area and if one can see who follows who and the venues they check-in on Foursquare and if Foursquare is used at all in Amsterdam - dig deeper into the limitations of these sources	a) Twitter data for a specific area can easily be collected through the Twitter application program interface (API) ATTENTION: only 1% of Tweets are geocoded! [https://opensource.com/article/17/6/collecting-and-mapping-twitter-data-using-r] b) Twitter data networks (who follows who) can be collected using the GET followers/ids and GET friends/ids for one specific user. the result shows who follows the specific user (followers) and who the specific user follows (friends) [https://developer.twitter.com/en/docs/accounts-and-users/follow-search-get-users/api-reference/get-followers-ids] c) it is possible to mine Twitter data for Foursquare check-ins by adding a HEAD request. again only small fraction of tweets can be analysed this way (first need to be geotagged, then need to have unlocked access to shared Foursquare check-ins [https://stackoverflow.com/questions/8206849/how-to-crawl-foursquare-check-in-data]	YES
2) Cranshaw et al. (2012) — Studying the structure and composition of a city based on social media its residents generate	Identifying how areas are used and by whom, how they are characterized and what shapes them	Determine the characteristics, usage and users of an area (socio-economic characteristics and demographics as indicator for gentrification) based on Foursquare check-ins collected from public Twitter timeline	1. determine Livehood clusters (characteristics, usage, users of an area) through aligning the a) check-ins from the Twitter public timeline (users) with the b) venue information from Foursquare API (usage and characteristics of area). 2. cross-check these determined neighbourhood profiles through interviews	Generally aggregating behaviour of many indvies behaviour of minorities. a limited sample was used; how an area really feels and what its social added value is can only b determined through interviews 1. check-ins are usually generated by a demographic of your professionals between 25-35, owners of smartphones and urban residents (this can influence results)	1. if proven accessible above (F5) then no issues 2. not a big data method, but should be considered for validity of results and regarding limitation(s)	Wait on G5	See H5 and b) the venue information can be collected through the Foursquare application program interface (API) [https://developer.foursquare.com/docs/api/venues/categories]	YES
3) Moseley et al. (2015) — exploring demographic attributes of social media authors for purposes such as smart ad placement, security, predator behaviour etc	Identifying gender, age education level of individuals in a neighbourhood	Determine demographic attributes (gender, age, education) of individuals in an area (demographics as indicator for gentrification) based the linguistic pattern of their tweets	1. determine gender, age and education level by downloading tweet texts an analysing them for abbreviation features	a) limited scope, tweezers of a specific age group were analysed b) prior knowledge as to which abbreviations are the essential characteristics of the to be determined demographics is necessary	a) a limitation for this research but generally 'just' an artefact of the data source. need to be aware that only a certain demographic uses social media b) this is a study on its own and cannot be adopted for my research			NO
4) Boy & Uitermark (2015) — restructuring and refracturing og the city through social media using social media to present, perceive and navigate the urban landscape and communicate belonging	Identifying how the city is perceived, used and represented and uncover socio-spatial inequalities	Determine characteristics and usage of an area as well as socio-spatial inequalities (socio-economic characteristics as indicator for gentrification) through Instagram	1. determine in what way the city is used through analyzing Instagram photos based on content and representation of content 2. determine networks of users (through likes and comments on photographs 2.a) determine user communities (segmentation) through looking at most central accounts, their backgrounds and what images they circulate as well as proportion of post at a taggés location (spatial embeddedness) 3. determine inequality through distribution of likes and comments (stratification)	Generally limited scope, only public accounts with geo-tagged photos are used 1.) accessible but a lot of work to analyse each photo for content 2.) again possible as data access is given but huge amount of work, same for 2.a) + 3.	Assumptions can be proven with a more adequate data source, e.g. Airbnb data 4) and 5)			Could, but NO
5) Bendler et al (2015) — spotting popular areas in the city through tracing social media of its users with the purpose of better city planning, venue recommendation and investment	Identifying the attractiveness of an area	Determine the level of attractiveness of an area (gentrification through attracting new residents) based on the spatial and contextual characteristics of Twitter activity of its users	1. determine the popularity of an area through the share of Twitter messages in that area compared to all available Twitter messages in the whole study area 2. determine the activity in an area through the sum of Twitter messages with additional content (photo, video, linked places) divided by the total amount of Twitter messages in that area 3. determine attractiveness of an area by overlapping the results of 1. + 2. (high popularity and high activity make a place more atractive)	Generally the necessity of establishing a grid over the study area and analysing each grid cell is challenging; limited amount of tweets are reviewed because only 1% of tweets contain the exact geographical location	Data from 1.+2. is accessible so there is no issue	How to get data from Twitter API??	See H5	YES
6) Paldino et al. (n.d.) — unconventional way of how people experience the city using information from geo-tagged photographs. compiling spatial behaviour of residents and tourist on global and local scales (local more interesting for me)	Identifying tastes of individuals in the city, what attracts them to live in a particular city or spend vacation time there	Determine the level of attractiveness of an area (gentrification through new residents and tourists) through photographs taken from Flickr	1. determine attractiveness of areas based on spatial distribution of number of photos taken (higher picture density means more attractive) distinguish between attractiveness for residents and for tourists they take pictures in different areas); a) determine who took the picture through their profile or amount of posts per user at a location over a given amount of time	Generally limited through the necessity of determining residents vs tourists, all photographs where no origin can be determined are excluded	1. the data from Flickr is openly accessible. but if Flickr the right platform to use in Amsterdam? / maybe Instagram is better?	Is Flickr data appropriate in AMS? If no is Instagram a good substitute?	Yes Flickr is used in Amsterdam (https://www.flickr.com/photos/walkingflatbumen/72157627140310742) Through Flickr API extract density of pictures taken and by whom	YES
7) Sobolevsky et al. (2015)								

	Focus	Goal	Method / Scenarios	Limitations	Data / Added Value	Tasks	Next Step	
Using digital information produced by human activity to examine human behaviour	Identifying city attractiveness for foreign visitors	**Determine level of attractiveness of an area (gentrification through tourists)** through barcard transactions, **geo-tagged photos (Flickr) and Twitter**	**1. Determine level of attractiveness** to tourists through amount if geo-tagged photos from Flickr (distinguish between tourist and residents) through examining where user took the longest time = resident) 2. Determine **level of attractiveness** to tourists through amount of geo-tagged messages from Twitter (again distinguish residents and tourists) 3. Determine **level of attractivenss** through bank card transactions of foreign banks	Limitations same as E20	**1. the data from Flickr is openly accessible. but if Flickr the right platform to use in Amsterdam? maybe Instagram is better? 2. data is accessible**	Wait for G20	See H20 and H5	YES
8) Saliman et al. (2017) analysing users's Twitter mobility patterns to determine key locations of individuals which give a higher resolution on urban land use than collective tweet analysis	Identifying key locations and how the urban landscape is used	**Determine usage of urban landscape** through analysing 39 million geo-located tweets, travel surveys and parcel-level land use maps	1. Random Walker Scenario: geo-located tweet of a user group seem randomly located over the city scape when viewed over time (tourist) 2. Preferential Return Scenario: geo-located tweets of users cluster around key locations when viewed over a longer time (e.g. home or work) 3. Semantic Coherence Scenario (results show this is not so true): if preferential return scenario is the case, the key locations can also disclose the functions. The implicit assumtion is made that the top key location for users' tweets is home 4. Temporal Coherence Scenario: it is also possible to assert each key location with the time of day based on the hours when the majority of tweets were posted. Assumambly, the timing of tweets is dependent on the land use type. Classification algorithms could be applied to infer land use types of key locations based on the timing of tweets at each location	Limitation of consistent data generation process	looking at individuals tweeted-from locations and their distribution within the parcel-level map is possible			YES
9) Abitol et al. (2018) how key linguistic variables measured in individual Twitter streams depend on factors like socioeconomic status, location, time, and the social network of individuals	Identifying socioeconomic characteristics of tweeters through analysing linguistic variables in their tweets	**Determine socioeconomic status of individuals from the way they tweet**	1. Collect a large Twitter corpus including geotagged posts and proxy social interactions of millions of users 2. Establish a detailed socioeconomic map describing average socioeconomic indicators with a high spatial resolution in the study area (here France) 3. Study 3 linguistic variables extracted from individual Twitter timelines were then studied as a function of the rate of standard negation, the rate of plural agreement and the size of vocabulary set 4. Process the information via a detailed multidimensional correlation study	Twitter provides a biased sample in terms of age and socioeconomic status as older or poorer people might not have access to this technology; sources of bias include the "homogenization" performed by INSEE to ensure privacy rights are upheld as well as the proxies we devised to approximate users' home location and social network; pre-knowledge of the language in which one is analysing linguistic paterns is a requirement				NO
SUMMARY	**Conclusion Assumptions**	Based on assumptions it is concluded, that Social MediaData can help determine whether a neighbourhood is gentrifying based on the level of social cohesion, socio-economic characteristics, usage and users in an area and the attractiveness of an area for residents and for tourist		status as older or poorer pe	**Added Value in my research**: Use Twitter data (Twitter social network and Foursquare check-ins) to determine A) social cohesion of an area B) socio-economic characteristics of an area C) the attractiveness through the popularity and the activity of an area (based on amount of Twitter messages and added content) Use Flickr data to determine attractiveness of an area through amount of geotagged photos (distinguish between locals and tourists)	**General Tasks**: Go in depth about detailed limitations of Twitter data (this includes Foursquare and Flickr data	**Next Step:** Explore if assumptions are valid and whether an added value is achieved	

Additional summary notes: Twitter: only a fraction of tweets is available through API access Foursquare: data needs to be mined through Twitter Flickr: ? Generally social media is only used by a fraction of the population, is this a problem?

Van: Maren Skinner [mailto:M.I.Skinner@student.tudelft.nl]
Verzonden: 2018
Aan: Housing Corporations in Nieuw West
Onderwerp: University Research: Big data in urban planning and policy matters

Dear sir or madame,

I am a master student at TU Delft, currently writing my final thesis, which concerns the exploration of possible advantages of big data in urban planning and policy matters. More closely I am exploring, whether big data offers the opportunity to better understand social structures of residents within a neighbourhood, through collecting information about them in unconventional ways. As a case study I have chosen Amsterdam Nieuw-West, as this area has been undergoing strong neighbourhood changes in the last years, reflecting interesting resident dynamics.

I am reaching out to the housing associations with property ownership in the Amsterdam Nieuw-West neighbourhoods Geuzenveld-Slotermeer, Slotervaart and Osdorp. I would like to know what kind of resident/tenant information is currently available to you, e.g. which information variables, such as socioeconomic factors income, education level, household size, etc. you value as relevant. Next, if it is possible for you to collect this kind of information easily. Secondly, I would like to know where you see a lack of information and what would be valuable information variables you would add to the information collection. Knowing this, I will then be able to determine in what areas there is an information gap, which big data could possibly bridge. I am hoping my research will result in an approach that will be able to support the decision making process for policy makers as well as planners to achieve outcomes that take intrinsic social structures of neighbourhoods into account, while achieving economic growth.
On this behalf, I am also interested in knowing, whether you see potential in big data as a tool for empowering housing associations in the decision making process of such planning policies.

I have attached a document to this email containing a list of possible information variables. Please go through it and make a mark if the information variable is valuable to you and if you are able to collect it. Feel free to add any other variables that you might be collecting.
There is also space provided for you to state the variables that are missing in your collection and possibly can be collected through big data. Last but not least, a field is provided for you to formulate your thoughts on the advantages of big data for housing associations.

Your effort will be of great value to my research. Please be assured that all collected information will be treated with care and used anonymously to draw general conclusions.

Please let me know, if you are interested in the research results. If so, I will be able to provide you with the final version of my report in July 2018.

Thank you very much for your time.

Kind regards,
Maren Skinner

Maren I. Skinner
Msc Student TU Delft
Management in the Built Environment
T: +49 (0)174 3206629
E: maren.skinner@gmail.com

Appendix B.II. Survey

Please fill in accordingly:

Information Variables	Yes, we value this information	Yes, we can collect this information
Income		
Education Level		
Education Status		
Employment Status		
Tenant Age		
Tenant Gender		
Ethnicity		
Marital Status		
Household Size		
Wish of tenant to move		
Wish of tenant to stay		
Satisfaction with the accommodation		
Satisfaction with the area		
Travel time to professional/educational facilities (work, school, university)		

Information Variables that should be added to the collection:

Thoughts on social media data as an empowering tool for housing associations:

	Geuzenveld-Slotermeer	Slotervaart	Osdorp	Overtoomse Veld	Contact details	Date of contact	Aknowledgement of Receipt	Follow up Email	Date of response	filled in Variables Table	Big Data response	Comment
de Alliantie	x	x	x	x	info@de-alliantie.nl	07.03.2018	Y	29.03.2018				
de Key	x	x	x	x	contact form	16.03.2018	Y	-				
DUWO	-	x	-	-	business@duwo.nl	07.03.2018	N	29.03.2018				
Eigen Haard	x	x	x	x	info@eigenhaard.nl	07.03.2018	Y	29.03.2018				
Rochdale; Gerard van Bortel	x	x	x	x	contact form; g.a.vanbortel@tudelft.nl	16.03.2018; 29.03.2018	Y	-	22.03.2018	Y	Y	interested in my final research results, as they are building a data set of their own
Stadgenoot	x	-	x	-	info@stadgenoot.nl	07.03.2018	Y	29.03.2018	1st response: 13.03.2018	N	-	not collecting any of the proposed variables; sent a followup email 29.03. after their first response
Woonzorg Nederland	x	x	x	-	klant@woonzorg.nl	07.03.2018	N	29.03.2018				
Ymere; Vincent Gruis	x	x	x	-	contact form; v.h.gruis@tudelft.nl	16.03.2018; 29.03.2018	Y	-	1st response (by Vincent): 29.03.2018; 2nd response: 10.04.2018	Y	Y	Vincent: email was forwarded; interested in my final research results as they are interested in everything that can help them understand their tenants and people of Amsterdam better
Cordaan (Extraordinary Member)	x	x	x	x	info@cordaan.nl	07.03.2018	N	29.03.2018	09.05.2018	-	-	cannot participate in my study