The Construction of Social Realities:

A Critical Data Studies Approach Towards the Representation of Open Data in Four Dutch Web Applications

Alex Smit (5950880)

Master Thesis

16-07-2018

Supervisor: dr. Mirko Tobias Schäfer

Second Reader: dr. René Glas

Master New Media & Digital Culture - Utrecht University

The Construction of Social Realities: A Critical Data Studies Approach Towards the Representation of Open Data in Four Dutch Web Applications

Abstract:

The Geluksmeter, Leefbaarometer, Uitkeringsmonitor and Benikgemiddeld web applications are examined through a Critical Data Studies framework to uncover their socio-technical apparatus and ask how they use and represent open data within graphic designs. The process of quantification of societal phenomena into data-driven infrastructures is depicted in which the four applications are understood as data assemblages (see Kitchin and Lauriault 2014). With an infrastructural inversion and discursive interface analyses the data assemblages are unpacked and examined. Through this methodological frame a critical analysis of the workings, affordances and implications of the four data-driven applications is shown. The argument is made to understand these applications, not as objective neutral accounting and representational tools of societal phenomena, but as interpretative non-neutral systems with a bias towards socio-cultural norms. Four key points are given how to improve data practices and the reading of graphic designs with the need for data visualization literacy when enacting such web applications which use open data to represent societal norms and values with graphic means.

Keywords: Data-driven web applications, data assemblage, Critical Data Studies, Leefbaarometer, Geluksmeter, BenIkGemiddeld, Uitkeringsmonitor

Alex Smit (5950880)

Master New Media & Digital Culture

Utrecht University

Index

Appendix

1. Introduction

In 2007, the Dutch right-wing platform *Geenstijl* constructed a web application called the *Allochtoon-o-meter.* This application appropriated statistical data of the Dutch Central Bureau of Statistics to visualize a measure on *'allochtones'*[1] residing within a zip code area in the Netherlands. The application was subject of heated debates as it singled out ethnic minorities, whereas *Geenstijl* answered that they simply used available open data to provide a public service (NRC 2009). This example shows how seemingly *neutral* statistical data can be appropriated by data-driven technologies for ideological means. As such, this thesis focuses on how four data-driven web applications use open data to represent and visualize socio-cultural norms and values. This examination originates from the field of Media and Culture Studies, in which socio-cultural norms and values represented in film, audio, text, and other sorts of media outings are primary point of study (Allen and Berg 2014; Hassler-Forest 2014). As new variations of aesthetic media artefacts and knowledge production entities are emerging with the rise of data-driven technologies and their ontologies, the primary concern regards how these technologies represent and affect epistemological conceptions of quantified societal phenomena.

With the omnipresence of data-driven technologies in society, questions concerning accountability for algorithmic bias, representation of socio-cultural norms and values, and appropriation of data gain attention in both academic and public discourse (Draude et al. 2017 p. 1-3; Iliadis and Russo 2016 p. 2-4; Kitchin and Lauriault 2014, p. 1-2; Ruppert 2015 p. 127-128). What is generally conceived as a neutral utterance of these technologies, i.e. the measurement and scoring of liveability within an area with the use of statistical data, actually consists of multi-layered contingent infrastructures making use of vast datasets, classifications, indicators and visualizations. Such data-driven technologies are part of an epistemic system in which complex statistical numbers are transformed into interactive graphic form and appropriated by various actors and agents. The knowledge produced through these systems is called into question when it turns out to prioritize socio-cultural norms and embed biases, while claiming to give objective representations of used data. As such, this research answers calls residing from the field of Critical Data Studies (from now on CDS), to regard such data-driven systems as *data assemblages* (Kitchin and Lauriault 2014, p. 5), and accordingly to scrutinize their infrastructures, arrangements, how data is generated and processed, the analytics used to gain insights and knowledge, their socio-political framing, and how they are employed (Dalton and Thatcher 2014; Kitchin and Lauriault 2014, p. 5-6).

The perception of data-driven technologies as *data assemblages* is used extensively in the

---

[1] Allochtones is a Dutch term regarding immigrants and non-native Dutch citizens of which one or more parents originates from another country (CBS 2017b).

project 'The Programmable City'[2], initiated by Rob Kitchin et al. (2017), whereas the use, deployment and procurement of data in urban dashboards, city management and other smart city technologies is examined. Understanding these technologies as *data assemblages* has proven to be useful to critically analyze how a society is translated into complex and versatile data-infrastructures, which are entangled within technological, social, political and economic apparatuses that affect and frame the generation, circulation and deployment of data (Kitchin and Lauriault 2014, p. 5-6; Iliadis and Russo 2016, p. 2).

Data residing from these *data assemblages* rarely appears "raw", and if it does, it is mostly a materialized, fabricated and contingent achievement (Gitelman 2013). Data thus needs to be apprehended as part of wider assemblages (Kitchin and Lauriault 2014, p. 6). In this context the following four web applications are regarded as *data assemblages*: *Geluksmeter* (from now on GM), *Leefbaarometer* (from now on LM), *Benikgemiddeld* (from now on BIG) and *Uitkeringsmonitor* (from now on UM). The GM, BIG and UM are constructed by the Central Bureau of Statistics (from now on CBS) as part of a broader open government project developed in 2014, in order to make statistical census data from the CBS accessible and more transparent to a wider public (CBS 2017). The LM[3] is constructed in order of the Ministry of Internal Affairs in 2008, in collaboration with the CBS, Atlas[4] and Rigo[5] to map, score and visualize the (average) level of liveability within the Netherlands based on one hundred indicators (Leidelmeijer et al. 2017).

The GM[6] is a web application designed to calculate and visualize a personalized score related to happiness or well-being of an individual by answering eight questions constructed upon eight dimensions: present financial situation, future financial situation, health, leisure, social life, government, safety, and living area (translated by author). The BIG[7] web application calculates and visualizes national averages based on certain criteria according to specified personas. The user can test if it conforms to the national average within five dimensions: culture and leisure, population, expenditures, health, and income (translated by author). The UM[8] is a web application which calculates the averages of six kinds of social welfare/benefits recipients within specified municipal areas, and visualizes it on a geographical map of the Netherlands. These data-driven web

---

[2] Within the Programmable City project sixteen researchers over the course of six years have worked on how "cities are increasingly being translated into code and data, and how these code and data are being used to transform how we understand, manage, work, and live in the city to produce smart cities" (Kitchin 2018).

[3] The Leefbaarometer application can be accessed via https://www.leefbaarometer.nl

[4] Atlas is an organization affiliated with municipalities within the Netherlands which document, analyze and process open government data for research purposes.

[5] RIGO is a similar organization as Atlas, but performs research not only for municipalities but for other institutional actors as well.

[6] The GM application can be accessed via https://www.visualisatie.cbs.nl/geluksmeter

[7] The BIG application can be accessed via https://visualisatie.cbs.nl/nl-NL/visualisation/benikgemiddeld

[8] The UM application can be accessed via https://visualisatie.cbs.nl/nl-NL/visualisation/personenmeteenuitkeringpergemeente

applications have been chosen as they represent, reaffirm and visualize socio-cultural norms and values by translating open data, originating primarily from the Dutch Central Bureau of Statistics[9] (from now on CBS), into interactive graphical designs. Although the socio-cultural and political contexts of these applications differ, they have the same data sources in common, and represent socio-cultural norms and values in a similar manner using a web application interface.

The research questions are constructed upon the contextual frame of Media and Culture Studies, in which the use and representation of open data from the Dutch Central Bureau of Statistics within four web applications is examined.

The main research question is: how is open (CBS) data used to represent values of social life within the Geluksmeter, Leefbaarometer, Uitkeringsmonitor and Benikgemiddeld web applications? This leads to the following sub-questions: Which values are represented within the four examined web applications? How is open (CBS) data graphically designed within their interfaces? And how do the four web applications produce knowledge concerning societal norms and values?

These questions are answered with an analytical scope in which the inner workings of the four web applications as *data assemblages* are unpacked using the concept of an *infrastructural inversion* (Bowker and Star 1999, p. 33). This concept enables to shed light on the invisible architectural configurations residing within these data-infrastructures (Bowker and Star 1999, p. 33-34). By making the embedded configurations visible, a more substantial understanding can be achieved of how data-driven systems use, process and transform data, and how they steer and render the knowledge production through their constructions. This means interrogating these technologies with according configurations and arrangements, which by design tend to stay in the background. Bowker and Star (1999, p. 34) elaborate on this notion that the concept of an *infrastructural inversion* is used to: "recognize the depths of interdependence of technical networks and standards, on the one hand, and the real work of politics and knowledge production on the other" (ibid.). It functions as a dual examination of data-driven technologies, not only focusing on the technical arrangements, but simultaneously examining the politics and epistemic notions embedded within the design. As in my opinion this approach neglects the user centred focus, the *environment of expected use* is examined with a discursive interface analysis according to 'The Walkthrough Method' (Light et al. 2016, p. 2-3). Through this analysis I show how the concepts of a *data assemblage,* together with an *infrastructural inversion* and 'The Walkthrough Method' can be applied to develop a better understanding of the internal logics and implications of data-driven web applications.

---

[9] The Central Bureau of Statistics is an administrative organization functioning autonomously from the Dutch Government. It "performs public service tasks, but operates independently and not under the direct authority of a Dutch ministry" (CBS 2017a).

I start elaborating with the interpretative nature of quantification practices and the epistemic notions of the use of open data in data-driven technologies. Next, the perspective of CDS on data-infrastructures is elaborated, whereas the concept of a *data assemblage* is operationalized. Within the analysis I examine the four web applications with a focus on the data-sources, data-practices, data-visualizations and the environment of expected use. Finally, the findings are discussed with a critical perspective on the *neutrality* and *objectivity* of open data, the representation and reproduction of socio-cultural norms, and four recommendations are given for a better understanding and awareness regarding the use and representation of data, and the implications of such data-driven technologies. I conclude with my argument statement to reconceptualise these applications, not as objective accounting tools of societal phenomena, but as interpretative systems enforcing socio-cultural norms and embedding them within the design of the web applications. Concluding, the findings are contextualized within a broader academic discussion taking place within the field of CDS, critically regarding data-driven technologies and their ontologies (Kitchin and Lauriault 2014; Kitchin 2015; Iliadis and Russo 2016).

2. A critical inquiry of quantification and open data practices

Quantifying society is a complex practice. Critics of quantification practices within the social, humanistic and natural sciences have often argued that reliance on complex numbers neglects the contextual and implicit aspects of those numbers (see Desrosiéres 2013 p. 329-330; Porter 1995, p. 3-6; Rottenburg and Merry 2015, p. 1-2). These numbers are not only used to measure, account and rate various societal phenomena, but are increasingly deployed to predict behaviour, sentiment, opinion and other personal characteristics. The appeal of numbers is especially attractive to institutional actors, as bias and arbitrariness are the most common aspects upon which these actors are criticized (Rottenburg and Merry 2015, p. 1-2). Policy constructed upon statistical data has an image of being *fair* and *neutral*, and thus facilitates a frame for governmental virtues such as justice and impartiality. Much of the existing data-driven systems appear merely as interfaces with lists of data, classifications, labels, all embedded within vast menu's and visualizations. These systems make use of various quantification practices to reflect on, and represent certain societal phenomena. These practices neglect the issue of quantification being a product of sequences of interpretive decisions and assumptions on how and what to quantify, categorize, code and label. As a phenomenon is more complex, it is harder to quantify, and the decisions made within the process become more difficult (Porter 2015, p. 35-36). These processes are bound by the availability of data, expertise, resources and purpose, which are the product of what actors have decided to collect, quantify, analyze, process, and in what to invest their resources (ibid.). The decision what to quantify

often resides in the concern what problem has to be examined. Quantification can enable getting an insight into a problem, as opaque relations can become visible. The core principle of quantification systems resides in the enablement of comparison and a common measure or divisor of data, in which I mean a certain similarity so that various data can be sorted, grouped, counted and calculated. When this principle of similarity is achieved, data can be translated into usable input for data-driven systems (Rottenburg and Merry 2015, p. 3-4). Sociologist Alain Desrosiéres (1998, p. 10-11) elaborates on this process as: "it allows a large number of events to be recorded and summarised according to standard norms". He argues that achieving equivalence is foremost a practical endeavour before it becomes a cognitive one. In this sense creating equivalences across data is a pragmatic practice, as random and unforeseen personal behaviour is translated into numbers, averages and patterns, and is consequentially transformed into statistical summaries of these different acts (Rottenburg and Merry 2015, p. 12-13). The awareness regarding this being a pragmatic practice helps the analysis of data-driven technologies in the sense that the seemingly *neutral* and *objective* used data can be questioned.

Census data was historically often accompanied by certain terms as 'index' or 'barometer', which are also used in the corpus, requiring that the used data should stand for something larger than themselves (Desrosiéres 2013, p. 329 - 331). Such statistical means are given credit for the construction of entities as unemployment- or crime-rates, and are often regarded as an effective measurement of government policy. Institutions do not generally characterize these numbers merely as interpretative indicators, but as measures of reality (ibid.). In 1978 Michel Foucault (2007, p. 387-389) had already described that quantitative instruments are historically not only used as tools providing evidence for certain ideals or arguments, but as instruments of governing, or as Foucault conceptualized: *governmentality*[10]. Within this line identifying, defining, categorizing and constructing quantified entities of constructed values as happiness cannot be perceived merely as neutral means of capturing, recording, storing and processing, but are related in an institutional and political manner (ibid.)[11]. Relatively new to this context is the construction of interactive web applications, which transform data into interfaces with interactive customizable visualizations, accessible by a broad public. The quantified entities are not only enacted by intelligible actors as it was primarily done in a historical sense, but by a broad public with various levels of data-literacy.

Through the means of quantification and measurement society becomes knowable, and this knowledge is structured, ordered and categorized with computational means (Desrosiéres 2013, p. 330-333). Numeric representation provides the generation of indexes and rankings of societal

---

[10] For Foucault (1991) *governmentality* means the mentality or art to govern, the emergence of professional and specialized forms to govern populations and territories.
[11] Sociologists and social theory scholars Pierre Lascoumes and Patrick Le Gales suggested the same argument in their book 'Governing Through Instruments' (2007).

phenomena, in which data is structured and simplified, where after it is transformed into visual form[12]. Once these indexes, rankings and visualizations are accustomed and naturalized within public debate, a tendency exists in which the normative assumptions, contingencies and messiness embedded within these formats is mystified (Ruppert 2015, p. 130-131). The quantitative reasoning and construction of indicators alongside these indexes and rankings have become points of argument surrounding objectivity in governance, open data and public statistics (Kitchin and Lauriault 2014, p. 2-3; Porter 2015, p. 34-36). Governmental actors are one of the key facilitators of public statistics and open data, as vast amounts of governmental produced and stored data are made openly accessible through i.e. web applications. The open data and open government movements frame data as if they are public goods and should be freely accessible and usable for all publics (CBS 2017a). One of the fields in which these points have been critically discussed is the field of CDS.

2.1 A Critical Data Studies approach

The interpretative and contextual aspects of data is one of the key focus points within the field of CDS. This emerging field examines the role and affects of data in society and claim to focus on "the types of research that interrogate all forms of potentially depoliticized data science and to track the ways in which data are generated and curated, and how they permeate and exert power on all forms of life" (Iliadis and Russo 2016, p. 2). Dalton and Thatcher (2014) conceptualized this frame as CDS, in which they urge to research and rethink data within the perspective of critical social theory, to examine the manners in which data can never be regarded as "objective, neutral, independent, raw representations of the world, but are situated contingent, relational, contextual, and do active work in the world" (Dalton et al. 2016, p. 1-2). Within the field of CDS it is argued that data generally are understood as being pre-factual, and pre-analytical, what is meant that it exists prior to argument and interpretation (Kitchin and Lauriault 2014, p. 3-4). Data should rather be understood as being constitutive of "the ideas, techniques, technologies, people, systems and contexts that conceive, produce, process, manage, and analyze them ... embedded within a larger institutional landscape of researchers, institutions and corporations, and are subject to socio-technical regimes" (ibid.). Reconceptualising data, as a given, towards *capta,* actively taken could enable the understanding of the constitutive foundation of data (Drucker 2011, p. 2). The notion of *capta* encompasses the understanding of observations as *taken and constructed* (from the Latin *capere*), instead of being *given* (from the Latin *dare*) (ibid.; Kitchin 2014, p. 2-3). The flexible, partial, subjective, and constitutive foundations of data should thus be apprehended within the context of *capta,* actively taken from a much broader and larger set of data infrastructures. Databanks, repositories, interfaces and algorithms can thus be regarded as expressions and exertions of *knowledge-power*: shaping how,

---

[12] For an example of this transformative process of data see appendix 1 on page 56.

and what questions can be asked, which answers could be given, how these are deployed, and most importantly, who should and could ask these questions (Ruppert 2015, p. 131-133; Kitchin 2014 ,p. 3).

The definitions generally used to describe and interpret data imply that they are merely part of technical process such as collecting, comparing, mining, storing, analyzing and processing (Kitchin 2014, p. 2). This connotes that only the uses of data are political, but not the data themselves (Kitchin and Lauriault 2014, p. 3-4). Grounded within Science and Technology Studies, CDS examines the flow of data by arguing that "the used data is simultaneously being produced by, and inextricable from software, hardware, instrumentation, protocols and documentation that helped to produce it" (Neff et al. 2017, p. 90). This means that data are the products of the *socio-materialities* of their production, and within this perspective data-driven technologies entail *data ecologies* (Vertesi and Dourish 2011, p. 533-534), *data assemblages* (Kitchin and Lauriault 2014, p. 5) and/or *knowledge infrastructures* (Edwards et al. 2013, p. 7-10)[13]. It could thus be stated that these technologies *mediate* data, or are *mediums* for the data (ibid.). Within this perspective the UM, LM, BIG and UM are understood as "data assemblages", which consist of:

the technological, political, social and economic apparatuses and elements that constitutes and frames the generation, circulation and deployment of data ... the apparatus and elements of a data assemblage may be systems of thought, forms of knowledge, finance, political economy, governmentalities and legalities, materiality's and infrastructures, practices, organizations and institutions, subjectivities and communities, places, and the marketplace where data are constituted ... more than the data system/infrastructure itself, such as a big data system, an open data repository, or a data archive, to include all of the technological, political, social and economic apparatuses that frames their nature, operation and work (Kitchin and Lauriault 2014, p. 5 - 6). [14]

---

[13] This is derived from Bruno Latour's Actor-Network Theory (ANT), as it relates to the concept of networks as socio-technological ensembles. ANT focuses on the inter-relational materialistic and semiotic elements of objects/subjects, actors/actants (Latour 2005). Within the vocabulary of Latour data-driven applications cannot be interpreted as a neutral intermediate which collects, calculates, and visualizes statistical data without steering and influencing the transference of knowledge embedded within these processes.

[14] This notion of a *data assemblage* is compared to Foucault's (1977) concept of the *dispositif*, which refers to a "heterogeneous ensemble consisting of discourses, institutions, architectural forms, regulatory decisions, laws, administrative measures, scientific statements, philosophical, moral and philanthropic propositions" (Gordon 1980, p. 194). A *data assemblage* could be perceived as a socio- technological variant of this *dispositif,* and within my interpretation it guides this research to not only examine the technological configurations residing within the data-driven applications, but widens my analytical scope to include the related political and socio-cultural apparatuses, which in turn can be understood as variations of *governmentality* (Foucault 1991).

The use of this approach is exemplified in the 'Programmable City' project, where i.e. Kitchin et al. (2015) examined the *governmentality* and logics of urban informatics and smart city technologies like the Dublin Dashboard[15]. They analyzed how this system used, deployed and mediated urban data, while simultaneously focusing on the wider socio-technological apparatus. By approaching this dashboard within the frame of CDS and a *data assemblage,* they were able to find that not one form of *governmentality* was being enacted by the Dublin Dashboard, but multiple as it inherits mutable contingent, abstract, dynamic and entangled logics which are translated and operationalized in context-dependent manners (ibid.). They argued that urban *governmentality* is transitioning from "subjectification – molding subjects and restricting actions – to modulation, modulating affects, desires and opinions, and inducing action within prescribed comportments ... as people become subject to constant modulation through software-mediated systems" (ibid, p. 1-3).

This modulation can be noticed within the GM, LM, BIG and UM as well, with the means of collecting, deploying and visualizing data with specific conventions embedded in political ideologies. The web applications modulate perceptions, opinions and frames of the public through their design and prescribed comportments, which affect the knowledge produced by these *data assemblages* (Kitchin and Lauriault 2014, p. 5). Statistical data is used, originating primarily from the CBS to measure, calculate and represent societal constructed values and virtues such as happiness and liveability, and make socio-cultural norms implicitly explicit (i.e. in the BIG application). The representation and appropriation of data within these systems cannot be perceived as objective and/or neutral, and the used data not raw but inherently "cooked", with recipes designed by cooks part of institutions embedding certain norms, aspirations and purposes situated within a broader epistemic frame (Gitelman 2013; Kitchin and Lauriault 2014). The *cooking* of data can be made explicit through the concept of the *looping effect*[16], coined by Ian Hacking (2007), and was inspired on Foucault's framework on the production of knowledge (Kitchin and Lauriault 2014, p. 8-9). Hacking elaborates that two converged processes are simultaneously at work within *data assemblages*, which "produce and legitimate its data and associated apparatuses/elements and shapes how its data do work in the world, that in turn influences future iterations of data and the constitution of the assemblage" (ibid., p. 8). The *looping effect is* one of those two processes[17], which regards the classification and organisation of data, how a data ontology comes into being, and how it

---

[15] The Dublin Dashboard is an open access, analytical dashboard which uses vast amounts of real-time urban data originating from smart city technologies to display information on traffic, weather, emergency's, public transport etc. (Kitchin et al. 2017).

[16] The five stages of a *looping effect* consist of: 1) classification; 2) objects of focus; 3) institutions; 4) knowledge; 5) experts (Hacking 2007).

[17] The second process is termed the "engines of discoverability", which do not only describe the 'making up' of people, but also of spaces (Hacking 2007). This second process perceives the work of a data assemblage while simultaneously legitimates and reproduces such work as the assemblage itself (Kitchin and Lauriault 2014).

can affect and reframe that what is classified (Hacking 2007, p. 287-289). Hacking (2007) argues that through the processes of the *looping effect* people can be *made up*, as is the case in the UM.

Within the UM this *looping effect* produces classifications regarding different kinds of social welfare/benefits recipients as people are reframed not as *regular* civilians, but singled out as *others* regarding recipients of social welfare/benefits[18] and visualized on a geographical map of the Netherlands. This is understood as a form of *profiling* and/or *social sorting*, which is currently the subject of much debate in the field of CDS (Kitchin et al. 2017; Iliadis and Russo 2016; Ruppert 2015). With *social sorting* certain groups in society might be prioritized above others, which can result in the exclusion of people and devaluation of their social status. Constructing such classifications implies the creation of a spectrum in which a progression is made from the individual to the general. These classifications often reflect the values, norms, and concerns within societies, while simultaneously affecting and shaping these subjects (Rottenburg and Merry 2015, p. 10-12).

A certain normative frame is used to construct these classifications. The concept of *standpoint theory* (Harding 2015) enables to examine such conceptions of normativity inscribed in data-driven technologies. The concept suggests that actors and agents regarding knowledge production practices should be attentive to power relations in which knowledge is always implicated in what perspective is used, who benefits from this perspective, and especially who does not (Draude et al. 2017, p. 1-3). From this *standpoint* a certain normative perception is made transparent, and acknowledges that what is left *unseen*, or *excluded* from the system. Creating such classifications and categories to enrich statistics and governance resides in a field of interpretation, influenced by reference frames, pre-existing categories and pragmatic concerns (Porter 2015, p. 34-35). Taxonomies are widely used to simplify these processes, but inherit a logic of grouping and labelling data, while they in practice do not exist together (ibid.). Forcefully grouping these data together can affect the meaning and contextual value of the initial data, as the initial data is related, layered, entangled and interacted. Bowker and Star (1999, p. 101 - 103) even go a step further, by arguing that the initial data used in such data-driven technologies form an ecology with compatibilities, and by breaking them apart and regrouping them within fixed categories and classification schemes alters their inherit meaning. Categorization and classification can thus be regarded as a divisive practice in which existing hierarchies and power structures are enforced (Foucault 2007, p. 96). Socio-political, cultural and technical experiences and reference frames influence the manner how these existing hierarchies and power structures are enforced (Rottenburg and Merry 2015, p. 20-21). I.e. data-

---

[18] The people are categorized and sorted into different kinds of social benefit and welfare allowances, upon which the indicators with according dimensions are constructed. The dimensions consist of: disabled (AO), disability allowance (AOW), work and income for the capability to work (WIA), work and aid to work for disabled young adults (Wajong), unemployment (WW), general welfare (Bijstand) (CBS 2016) (translated by author).

driven technologies make use of pre-existing data, which were initially collected for different purposes, as this is generally less expensive than the collection of 'new' data. Additionally, the amount of categories within these systems cannot precede a certain limit, as they need to fit in a highly structured format for people that have to work with them (Bowker and Star 1999, p. 102-104). This exemplifies that the coding processes are affected by socio-political contexts, not only by certain assumptions and taxonomies, but by infrastructural and institutional arrangements, purposes, intentions, interests, expertise and resources (Rottenburg and Merry 2015 p. 5-8).

Drawing from this context, within the LM application people are classified and then sorted according to one hundred indicators concerning different characteristics and values regarding liveability. These indicators are grouped, divided and transformed into five dimensions: security (24%), population (15%), facilities (25%), physical surroundings (18%) and buildings (18%) (translated by author). The value shown behind each dimension is the value each dimension has for the calculation of the liveability score, which shows how the hierarchy of the theme's for the liveability measure is structured. According to these dimensions a score on liveability is visualized on an interactive geographical map of the Netherlands with the help of nine colours[19], as seen in fig. 1. The map uses municipal boundaries whereas the user can navigate the map using the arrows on the right-side of the screen. The user can zoom in on a grid level of one hundred times one hundred square meters. In fig. 1 almost all municipalities are scored positively (except Rotterdam), it needs to be taken into account that these scores are in relation with the national average, which consists of a score of above average or "ruim voldoende" (Leefbaarometer 2017). This means that the negative score for Rotterdam indicates that the level of liveability is below the national average.
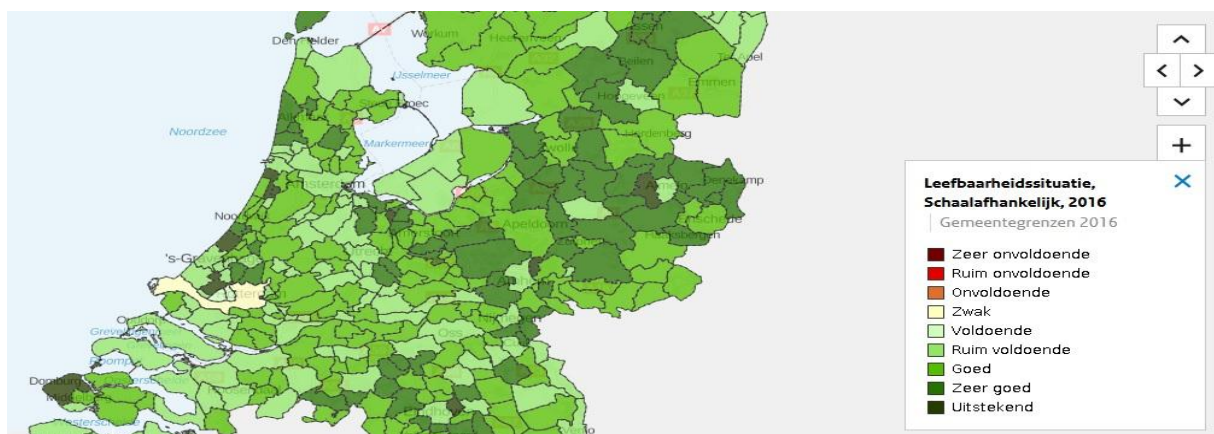

Figure 1: Leefbaarometer map[20]

---

[19] The nine possible colours score the liveability within a postal code area, differentiating from dark red (very insufficient) to dark green (excellent). There are three different shades of red (very insufficient to insufficient), one yellow (weak) and five shades of green (sufficient to excellent).
[20] Image acquired from https://leefbaarometer.nl/kaart/#kaart

Not only are people categorized and sorted into constructed indicators and dimensions, they are scored depending on certain normative assumptions and measures which fits into the hegemonic structure. Scoring citizens with data-driven infrastructures based on implicit norms translated into indicators is a problematic endeavour, as these systems cannot be understood as *neutral*, *objective* and all-inclusive. These applications seek to be transparent and represent data in an *objectiv*e manner, while simultaneously entrapping themselves in a paradoxical effect of suspicion from the public in which it seems they have something to hide by excluding data, or data which represents certain social groups as the 'native' Dutch citizen (Ruppert 2015, p. 145-146). As the LM embeds a strong focus on minority groups and their negative value for the liveability calculation, it includes certain dimensions of data and exclude others what can be understood as an example of the *standpoint theory* concept (Harding 2015). As such, the constructors of these applications prioritize the socio-cultural norms they value. As what is behind the data is forgotten, the choices made within the system are mystified and it is unknown what has been excluded and why, because it actively promotes certain realities and renders others opaque.

2.2 Imaging objectivity

Data visualizations and infographics are generally regarded as simplified ways to look at, and comprehend data. However, data visualizations and infographics are not the same tools to visually represent data. The difference lies within the function and understanding of the two, whereas data visualizations lets users draw their own conclusions based on their interpretations of the visualized data, and infographics show highly mediated information with the purpose to actively guide the users onto conclusions (Kennedy et al. 2018). As such, an infographic can potentially contain data visualizations, but data visualizations cannot contain infographics.

Both within data visualizations and infographics it is a misconception to regard the represented data as a given, as researchers (or other entities) actively decide what aspects of our lived experience they do, and do not record. This has a different connotation embedded within, and creates awareness regarding the interpretative dimension of data and their visualizations. Johanna Drucker (2011, p. 8) argues that "the graphical force conceals what the statistician knows very well — that no "data" pre-exist their parameterization". This "parameterization" affects the transference of knowledge by the visualisations. Helen Kennedy et al. (2016, p. 726) adds to this notion with the argument that certain conventions embedded in visualization practices function to give graphic design the framing as if they are objective representations of data: "conventions work to imbue visualisations with the quality of objectivity (which brings together other qualities like transparency, scientific-ness and facticity). This produces the impression that visualisations are showing the facts, telling it like it is, offering windows onto data" (ibid.). Kennedy et al. (2016) note how despite their

subjective nature, visualizations are often used to naturalize and objectify. The translation into graphical form gives it a certain simplicity and legibility, which could hide and mystify several aspects of the initial interpretative framework on which the original statistical data were created (Drucker 2011, p. 6-8). This easily readable visualization produces the impression that it is showing the facts, as if it is offering a clear window onto data.

Kennedy instigates to deepen the understanding of representation of data within visualization technologies by comprehending the conventions embedded within these practices. According to Kennedy the key points to look for consist of: data; key, axes and scales, title and the inclusion of data sources (Kennedy et al. 2018). These conventions, along with the power of graphical design, have to be taken into account when scrutinizing visualizations. Furthermore, the complexities and rhetorical capacities of visualizations have to be taken into account simultaneously while enacting data-driven technologies. As Aiello (2007, p. 154) argues "we mostly rely on visualized data's appearance to understand their substance", visualizations are vital in communication between - in this case – the CBS and the general public. By interacting with the interfaces of the applications and visualizations as non-neutral, the appearance of the visualizations can be tested according to its substance (Gitelman 2013, p. 11-12).

Visualization technologies have historically been deployed to minimize the mediation of the designer (Daston and Galison 2008, p. 664). However, while held up to have epistemic virtue, visual technologies commonly mediate objectivity. Daston and Galison (1992) examined this within their research on scientific atlases. They argue that objectivity has a history, and that there has to be a time before this objectivity has come into being (Daston and Galison 2008, p. 666-667). They show the constructed origin of this objectivity with the analysis of various scientific atlases[21]. As these atlases are resource intensive to remake - among dimensions of skill, time and money – they could be understood as indicators and barometers of shifts in epistemic values (ibid.). Consequently, a new atlas is a product of a community of scholars and practitioners to see the world anew. The scientific atlases are thus normative, they exemplify a perspective or specific way of seeing. As Daston and Galison (2008, p. 667) elaborate: "maps don't prescribe rules on how to see; they embody a set of practices. To read a map is to master a way of seeing by doing ... to use an atlas is to see, over and over again, and thereby to learn how to see in a specific way". This is an example of how conventions and frames get normalized, and how objectivity comes into being. Thus, objectivity is understood as a history of learning how to see, and visualization technologies mediate this objectivity with the use of conventions and frames (ibid.). The resemblance from this example to the applications can be made explicit when the focus is held on the use of geographical maps and conventions as

---

[21] Atlases were used because of their longitude as a scientific genre; their broadness across a wide array of empirical sciences; and their centrality in education and scientific practices (Daston and Galison 2008).

geometrical shapes and lines within the visualization practices of the applications. As Galloway (2011) argues in the case of the visual turn of data, visualization technologies are foremost depictions of principles and codes of production, and only secondarily of the initial data. The data assemblages are thus understood as "productive agents and distinctive multi-layered elements in the epistemic process" (Bredekamp et al. 2015, p. 2-3). The translation process of statistical data into graphical form, according to an example, is shown in appendix 1 on page 56.

3. Methodology: analyzing data assemblages and data visualizations

In this chapter I elaborate on how I unpack the data assemblages; how I analyze the graphic designs; how I examine the official documentation regarding the web applications; and how I scrutinize the environment of expected. First, I elaborate on the corpus, where after the methodological frame for the analysis is described. Finally, the limitations of this frame and research are depicted.

3.1 Corpus

The corpus consists of four web applications[22], official documentation regarding the applications and the graphical user interfaces. All four applications use Dutch open (government and/or census) data provided primarily by the CBS, to represent various societal phenomena on maps, graphs, charts and other graphic designs. The GM, BIG and UM are constructed by the CBS as part of a broader open government project developed in 2015, in order to make statistical data from the CBS accessible and more transparent to a wider public, and let users interact and experiment with the data (CBS 2014). These applications use data originating from 2010 to 2016, and were updated for the last time in 2016. The LM is constructed in order of the Ministry of Internal Affairs in 2008, in collaboration with the CBS, Atlas[23] and Rigo[24] to map, score and visualize the (average) level of liveability within the Netherlands (Leidelmeijer et al. 2017). The four web applications have been chosen for their similar data sources, visualization practices and purpose to inform, while differentiating in their means of representation and appropriation of data. Furthermore, the GM, LM and UM have official documentation regarding the workings, functionalities, motivations, purpose and context of the applications. As such, the following documents are incorporated within this research to examine the motivations and purposes of the applications: 'Welzijn in Nederland' (CBS 2016), 'Leefbaarometer' (CBS 2017); 'Leefbaarheid in balans' (Leidelmeijer et al. 2011); 'RIVM Rapport 630950002 Kwaliteit van de Leefomgeving En Leefbaarheid' (Leidelmeijer et al. 2008); and 'Arbeidsmarkt in Cijfers' (CBS

---

[22] Geluksmeter (GM), Leefbaarometer (LM), Ben Ik Gemiddeld (BIG) and Uitkeringsmonitor (UM)

[23] Atlas is an organization affiliated with municipalities within the Netherlands which document, analyze and process open government data for research purposes.

[24] RIGO is a similar organization as Atlas, but performs research not only for municipalities but for other institutional actors as well.

2016a). These documents are embedded in the research to gain a more nuanced and broadened conception of the applications. Special focus is held on how these are constructed, what purpose they serve, the motivation behind specific design choices, the framing and discursive aspects, affordances, and what data is used. As the LM has more documentation regarding the functioning and motivation of its workings, and uses more data sources than the other applications, the analysis of the LM will go more in-depth in relation with the analyses of the other three applications. Four different dimensions within the research corpus are isolated. These consist of: the statistical data functioning as input for the systems, the web applications and their interfaces, the official documentation accompanying the applications, and the environment of expected use. These are examined within the analysis in given order with the following approach.

3.2 Multidimensional Approach

A multidimensional approach is constructed in which the applications are understood and examined as *data-assemblages* (Kitchin and Lauriault 2014, p. 5), and their technical configurations unpacked with an *infrastructural inversion* (Bowker and Star 1999, 102). This approach is depicted in an overview in fig. 2. As no numbers are known on the usage of the applications, and the assembled approach lacks the focus on the imagined audiences and data publics of these applications, the 'Walkthrough Method: an approach to the study of apps' is used to examine the "environment of expected use*"* Light et al. (2016, p. 2-4). This method enables the uncovering of the "intended purpose, embedded cultural meanings, and implied ideal users and uses" of an application/interface (Light et al. 2016, p. 1). Light et al. (2016) present a method which enables to directly engage with an applications interface, in order to examine its technological mechanisms and embedded cultural references to understand how it guides users and shapes their experiences. 'The Walkthrough Method' involves the researcher examining what the application contributes to users' interactions with it. This more closely resembles components of Critical Techno-cultural Discourse Analysis (CTDA), which examines an artefact's interpellation, or calling out, of certain identities according to its embedded ideologies. Whereas CTDA examines a technology's interface alongside user practices, 'The Walkthrough Method' focuses on interface elements and their connection to the applications environment of expected use[25].

The researcher adopts an STS approach of systematically producing a collection of data by generating recordings, such as through screenshots, while conducting the walkthrough. This involves

---

[25] This means a step by step observation and documentation of screens, features and flows of activity, whereby actions and interactions become available for critical analysis. This process is contextualized within a review of the applications vision, operating model and governance, that will form an idea about the environment of expected use, which means how app designers anticipate it will be received, generate profit or other forms of benefit, and regulate user activity. For a broader description of this method see appendix 2 on page 58.

attention to the app's materiality and the physical interactions encouraged by the app, it involves the analysis of the affordances and actions it requires a user and user guides it presents, and other affordances like buttons, icons, and so forth, and finally it requires imaging how users would perceive these as affordances or constraints. This approach is grounded within STS and ANT, which relates to the framework of CDS.

To structure the analysis of the graphic designs following Helen Kennedy's (2018) 'Seeing Data' project, five features are used as a frame of perception: title; axes and scales; key; data sources; and dataset (Kennedy et al. 2018). First the focus is held on the presence and depiction of a title: if a title is present does it help you to better understand the visualization, and do the visualizations support the title? Second, the used axes and scales are considered: what scales are used, how are they used, what is their starting point, their range and value, which variations in values are visualized, and which are valued positive and which negative? Third, the key (legend) is examined: is there a key present and how does it inform you about the used colours, shapes and variables within the visualizations? Fourth, the focus is held on the data sources: are there data sources present, how are they presented, where does the data come from, and if no data source is presented how can the user be sure that the used data is accurate and trustworthy? And fifth, the used and represented dataset: how is the data related in the visualization, which categories are used and how are they ranked, which patterns are presented and what trends, and in what kind of graphic design means the data are presented (bar charts, line charts, pie charts, etc.) (ibid.). Within this structure the graphical designs of the applications are examined and placed in a schematic overview.

The following methodological overview is constructed to show this approach and isolated elements of the data assemblage (Kitchin and Lauriault 2014) as key points of study (see fig. 2).

What
- GM, LM, BIG, UM as research objects

How
- Data assemblage: use of data; representation of data; appropriation of data
- Infrastructural Inversion: technical configurations
- Graphical design: title; axes ; scales; key; data sources; and dataset
- Walkthrough Method: environment of expected use; vision
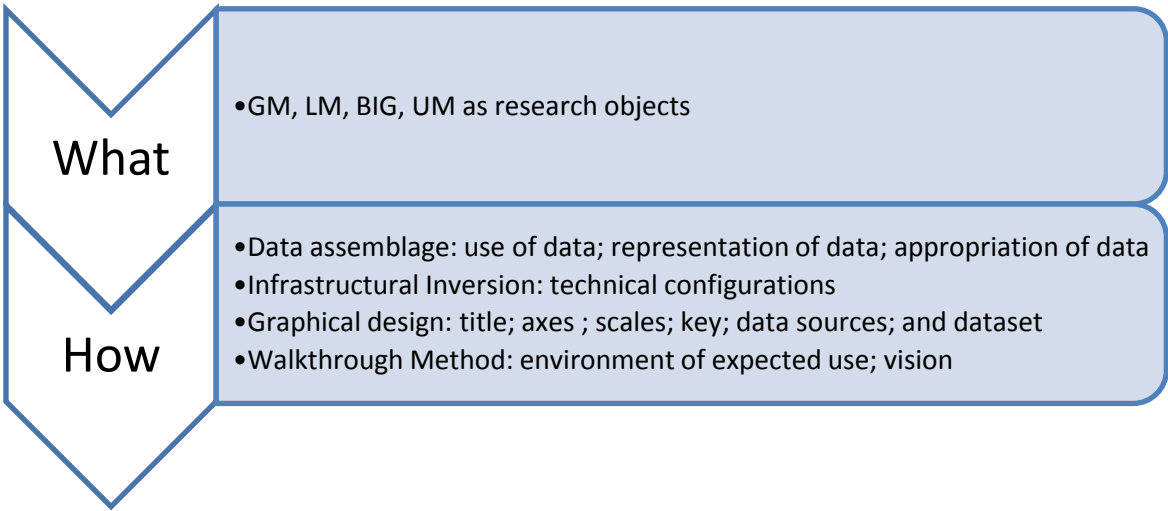
Figure 2: Methodological structure

3.3 Limitations

I find that traditional interface-centred approaches fall short in the complexity and representativeness of the used data and the imagined audiences are difficult to identify and portray. As such, one of the most important limitations within this research consists of the lack of insight into the audiences and publics of the four web applications. No numbers are known on the usage of these applications, how they are used, and why they are used. Consequently, it is unknown how the knowledge generated through these applications is appropriated by the users. 'The Walkthrough Method' enables to partially circumvent this limitation, as it allows to examine projected users based on for whom the applications are intended by design, and those who are not projected in the scenarios of the applications. Still, this is based on interpretations of official documentation and the design, which could project a somewhat interpretative frame of the imagined audiences. Furthermore, the GM, BIG and UM have not been updated since 2016, and run on data originating from 2010-2016. This means that their constructed visualizations do not consist of present-day societal phenomena, which has to be regarded critically when presenting the findings. The BIG application does not have official documentation regarding the construction and motivation of the system. This means that the analysis embeds a gap for the construction and motivation of configurations within this application.

Finally, it is not feasible to embed the full discursive interface analyses of the four web applications within the thesis. To avoid the effect of cherry-picking in the analysis, and give context to the findings, the discursive interface analyses as part of 'The Walkthrough Method' (Light et al. 2016) are shown in appendix 2 on page 58.

4. Unpacking data assemblages

First, the use of statistical data from the CBS by the four web applications is described. Next, the use of data visualizations and its conventions is analyzed. Additionally, the environment of expected use is examined and concluded with four points of critique and argumentation for data literacy. The analysis starts with a short description on the context and functioning of each application, hereafter the analysis goes in-depth. Due to length of the analyses and the limited space within the thesis, only the LM and GM are extensively examined within this chapter, as these applications have more complex interfaces and more substantial documentation regarding its functioning and motivation. Additionally, the full discursive interface analyses of the UM and BIG applications can be found in appendixes 2.3 and 2.4 on page 79-89. To outbalance the analysis of the applications more attention is paid to the graphical designs within the UM and BIG applications.

The LM application maps, scores and visualizes the (average) level of liveability within the Netherlands. The measure of liveability and the construction of the application are based upon the documents: 'Leefbaarheid in balans' (Leidelmeijer et al. 2011); 'RIVM Rapport 630950002 Kwaliteit van de Leefomgeving En Leefbaarheid' (Leidelmeijer et al. 2008). It calculates a liveability score based on one hundred indicators, divided into five dimensions: security (24%), population (15%), facilities (25%), physical surroundings (18%) and buildings (18%) (translated by author). The percentage behind each dimension depicts the value it inherits for the liveability measure, as visualized in fig. 3.
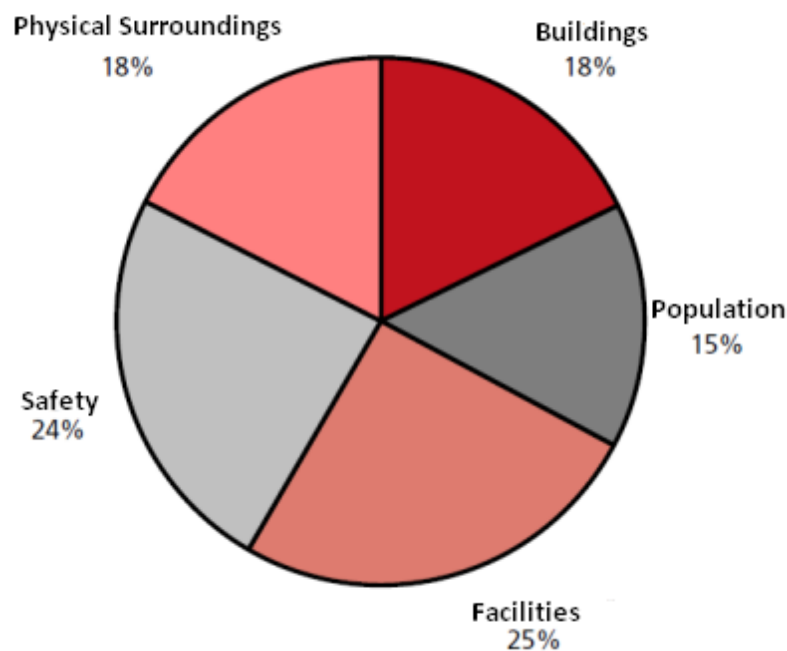
Figure 3: Value of dimensions in the liveability calculation

According to these five dimensions a score of liveability is visualized on an interactive geographical map of the Netherlands with the help of nine colours (see fig. 1 on page 14). The user can select different spatial dimensions which alter the visualized geographical map regarding a municipal scale, neighbourhood scale and a grid-map scale. The discursive interface analysis as part of the 'Walkthrough Method' (Light et al. 2016), in which the interface of the LM is 'walked through' can be found in appendix 2.2 on page 63.

In 2014 a second version of the LM (2.0) has been developed as the first version (1.0) was flawed regarding the value of the population dimension for the liveability calculation, based on new findings and held surveys attributing what factors people deemed important for a liveable environment (see CBS 2017a; Leidelmeijer et al. 2011; Leidelmijer et al. 2017). The second and current version of the LM is updated every two years with recent data and minor adjustments to the design and interface based on feedback given by users.

The GM[26] is a web application designed to calculate and visualize an average score related to happiness and/or well-being. It has been developed by the Centraal Bureau voor de Statistiek (CBS) in 2016, based on the report 'Welzijn in Nederland' (2016), with statistical data on the wellbeing or satisfaction of Dutch civilians in 2015 (CBS 2016). In this report several surveys are documented among the Dutch population to establish average scores within eight dimensions which relate to happiness. These dimensions consist of: present financial situation, future financial situation, health, leisure, social life, government, safety, and your living area (CBS 2016). In order to calculate your personal happiness score, eight questions have to be answered on a scale from one to ten, based on these dimensions. After answering these questions your personal happiness score is visualized within a circular graph, and can be compared with average scores relating to a specific demographic group which resembles your provided answers. This process is shown in appendix 2.1 on page 59 within the discursive interface analysis of the application. In fig. 4 an example is given of this circular bar chart specified with demographic variables as gender, age group and education level. The selected variables here are: male; 25-44; WO. Each coloured circle relates to one dimensions which becomes clear after the cursor is hovered over the graph. The average happiness score for my selected demographic group is 7,4, which the user can now further personalize by clicking on the circle and answering eight questions regarding your personal situation.
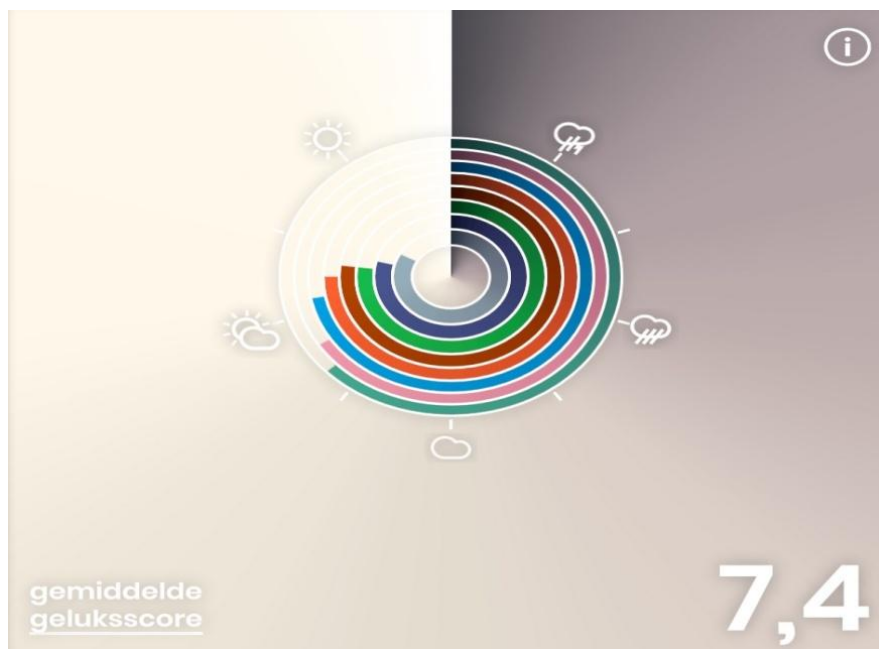
Figure 4: Geluksmeter circular bar chart [27]

[26] The GM application can be accessed via https://www.visualisatie.cbs.nl/geluksmeter
[27] This figure is a screenshot of the GM interface, acquired from https://www.visualisatie.cbs.nl/geluksmeter

The UM[28] is a web application which calculates the averages of six kinds of social welfare/benefits recipients within specified municipal areas, and visualizes it on a geographical map of the Netherlands (see fig. 6).
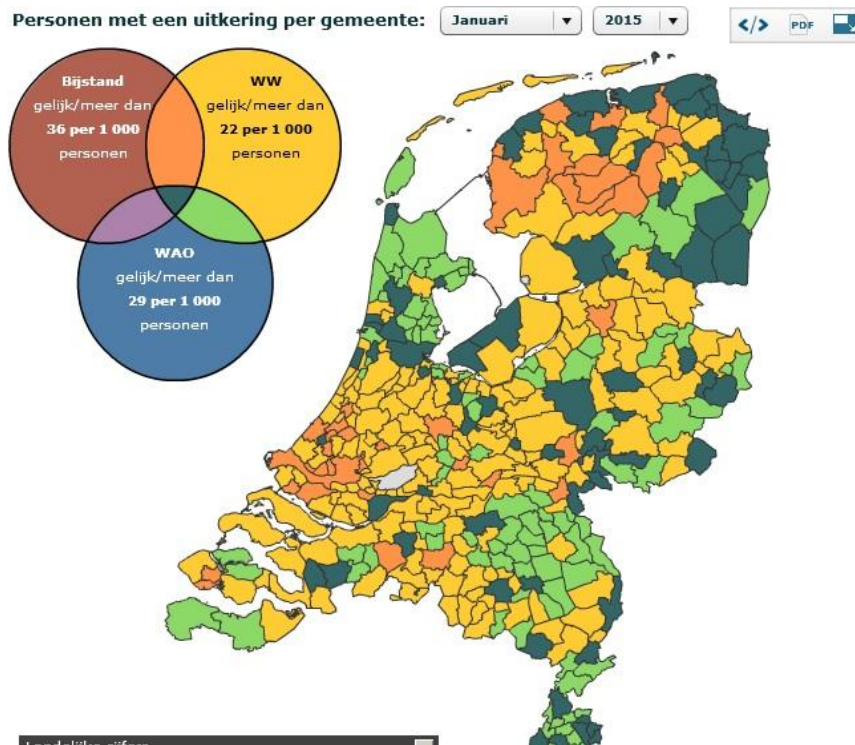
Figure 6: Uitkeringsmonitor map [29]

The user can select the following kinds of social welfare/benefits within a Venn diagram to be visualized on the map: law for the disabled (AO), law on disability allowance (AOW), law on work and income for the capability to work (WIA), law on work and aid to work for disabled young adults (Wajong), law on unemployment (WW), and the law on general welfare (Bijstand) (translated by author). A colour can then be assigned to each category of social welfare/benefits recipients whereas it is visualized with a comparing measure of social welfare/benefits recipients of one per thousand inhabitants within a specified area. In this example the user selected dark red for the general welfare recipients, yellow for the unemployment welfare recipients and blue for the disability allowance recipients. The Venn diagram mixes these colours to represent a mixture of these different kinds of welfare recipients on the geographical map of the Netherlands, on a municipal scale.   The discursive interface analysis of the application as part of the 'Walkthrough Method' (Light et al. 2016), in which the interface is 'walked through' can be seen in appendix 2.3 on page 78.

---

[28] The UM application can be accessed via https://visualisatie.cbs.nl/nl-NL/visualisation/personenmeteenuitkeringpergemeente
[29] This figure is a screenshot of the UM interface, acquired from https://visualisatie.cbs.nl/nl-NL/visualisation/personenmeteenuitkeringpergemeente

BIG[30] is a web application which calculates and visualizes the national averages based on certain criteria which the user selects within the interface. Based on these criteria the user can test if it conforms to the national average within five dimensions: culture and leisure, population, expenditures, health, and income (translated by author). Within these dimensions a sub-division is made in which the user can further personalize the variables. The personalization variables consists of the user being able to select a specific persona, age group and gender (see fig. 5).

Figure 5: Ben Ik Gemiddeld persona creation [31]

The user has to choose one of the visualized personas and is not able to create its own persona. The application then calculates and visualizes a national average based on the selected criteria, illustrated within various visualizations. The discursive interface analysis of the application as part of the 'Walkthrough Method' (Light et al. 2016), in which the interface is 'walked through' can be seen in appendix 2.4 on page 83.

4.1 The use, representation and appropriation of data

The CBS provides data freely on their website (https://www.cbs.nl), which the user can see, interact with and download with an Application Programming Interface (API). This API also facilitates access to CBS databanks consisting of assembled data residing from several databanks as: Bisnode, Rigo, Atlas, ABF Vastgoedmonitor, Kadaster, Municipal Administrations Office (GBA), Centre for Work and Income (CWI), National Police Corps (KLPD) and Policemonitor (CBS 2016). The datasets are accompanied with metadata, which provide contextual information on how the data has been collected, what measures and formats are used, in what timeframe, etc. This metadata accompanies the datasets and can be downloaded through the API as a separate .CSV file. See fig. 38 in appendix 2.2.1 on page 75 for an example of how metadata is shown through the API. The CBS developed this

---

[30] The BIG application can be accessed via https://visualisatie.cbs.nl/nl-NL/visualisation/benikgemiddeld
[31] This figure is a screenshot of the BIG interface, acquired from https://visualisatie.cbs.nl/nl-NL/visualisation/benikgemiddeld

service to increase governmental transparency and accessibility to open data on the Dutch state (CBS 2017b). This API is part of the Open Data Statline platform[32] (from now on ODS platform). The four web applications are part of this platform[33] and are a small selection out of a total of fourteen visualization applications incorporated on the ODS platform[34]. These fourteen applications are categorized (by author) as stand-alone web applications, as they have an own interface with according menu's and graphical means. What is noticeable is that the GM, BIG and UM use open data originating from 2010 to 2016. The LM is the only application using data from a broader time spectrum: 2002 to 2017.

4.1.1 The Leefbaarometer

The LM application uses data primarily from the CBS and the GBA, enriched with data from Bisnode, Rigo, Atlas, ABF Vastgoedmonitor, Kadaster, CWI, and the KLPD for the calculation of a liveability score[35]. This data contains statistical information on people's age, income levels, state of welfare, surroundings and the amount of social benefits in an area, information on criminal records, property ownership, finance, social benefits, municipal records, and work history. With the help of these data sources one hundred indicators are constructed, divided into five dimensions on which a statistical estimate of the liveability is presented. The LM calculates a measure of liveability within a postal code area depending on the presence and availability of data from various data sources. The dimensions are translated into the score depicted on the interactive geographical map of the Netherlands, with coloured zones and navigation possibilities (see fig. 7).

Figure 7: Postal code area map Leefbaarometer [36]

---

[32] The Open Data Statline platform can be accessed via https://opendata.cbs.nl/statline/portal.html
[33] The Leefbaarometer is the only web application not only accessible via the Open Data Statline platform, as it also has its own website: https://www.leefbaarometer.nl
[34] See appendix 3 on page 85 for the full list of visualization applications.
[35] For more information on how the used open data can be downloaded by the user, and what metadata is accompanied see appendix 2.2 on page 68.
[36] Image acquired from https://leefbaarometer.nl/kaart/#kaart

In fig. 7 the City of Utrecht on a neighbourhood scale is visualized. You can see that most neighbourhoods have a positive score, except Overvecht and Kanaleneiland. The user can navigate on the interactive map using the arrows on the right side of the screen, and can zoom in and/or out using the "+" and "-" buttons. Pressing on the circle symbol makes the map shift towards the GPS location of the user, if GPS is turned on. Pressing the "i" button provides an information screen where the user can find additional information on the liveability calculation and links towards instruction pages where more information on the application and its calculation models is provided.

Additionally, when clicking on a scored area, and scrolling down the visualized window of information, it becomes clear that the scores are comparisons to the national average of liveability (see fig. 8). The scores are divided into the five dimensions of the LM, with an additional bar chart showing the deviation regarding the total score of that area in comparison with the national average.
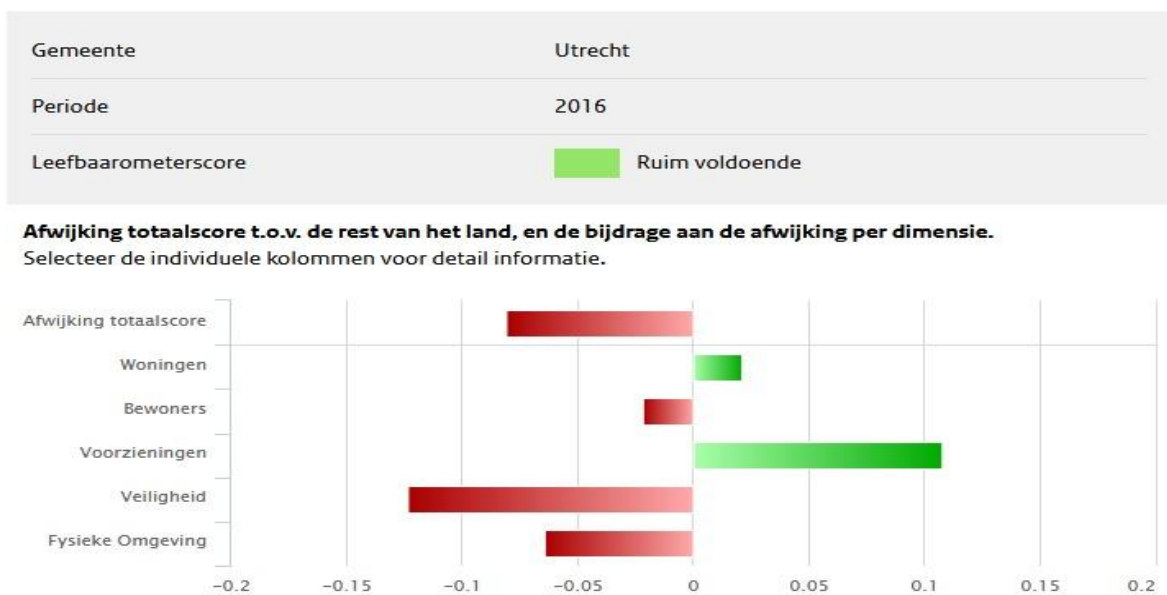
Figure 8: Comparison liveability scores in bar charts[37]

A scale from -0.2 to 0.2 is depicted, which means a spectrum from negative 20% to positive 20%, with 0 regarding the average. This scale is constructed as the provided scores all range within this spectrum. The colour red is used if the given scores are lower than the national average, and green is used as the scores are higher than the national average. The user can see how each individual dimension is scored in comparison with the national average. Additionally the user is forwarded to the official document regarding the construction of the LM and its statistical models (Leidelmeijer 2017).

An examination of how a liveability score is constructed shows how the statistical model focuses on the measure of various conditions of a habitat, which can affect the judgement on, and

---

[37] Image acquired from https://leefbaarometer.nl/kaart/#kaart

value of this direct environment of an individual. This is exemplified when a closer look is taken at the constructed indicators within the population dimension of the LM. The liveability score is for 15% determined by the population dimension. This dimension consists of sixteen indicators (see fig. 9), which are divided in five subcategories (by author): seven national origins indicators (one to seven), three household composition indicators (eight to ten), two economic indicators (eleven and twelve), one age indicator (thirteen), and three development indicators (fourteen to sixteen).

Nr	Derived	Data Source
1	Percentage of Western allochtones	CBS (GBA)
2	Percentage of non-Western allochtones	CBS (GBA)
3	Percentage of Middle- and Eastern-Europeans	CBS (GBA)
4	Percentage of Moroccans	CBS (GBA)
5	Percentage of Surinamese	CBS (GBA)
6	Percentage of Turks	CBS (GBA)
7	Other non-Western allochtones	CBS (GBA)
8	Percentage of one parent households	CBS (GBA)
9	Percentage of multiple person households (no children)	CBS (GBA)
10	Percentage of multiple person households (with children)	CBS (GBA)
11	Percentage of disability benefits recipients	CBS (GBA)
12	Percentage of social welfare recipients	CBS (GBA)
13	Percentage of elderly	CBS (GBA)
14	Development of persons within age group 15-24 years	CBS (GBA)
15	Development of households	CBS (GBA)
16	Number of mutations	CBS (GBA)

Figure 9: Indicators Table[38]

Indicators one to seven refer to the number of "allochtones" with various origins and ethnicities living in a specified zip code area. Indicators eight to ten refer to the composition of the households, incorporating the average number of people in a household, whether or not there are children in the house and if there are one or two parents in the household. Indicators eleven and twelve refer to the number of people that receive social benefits and/or welfare from the state, due to disabilities and unemployment. The indicator thirteen is the only indicator which focuses solely on one specific age group, which consists of elderly (65+). Indicators fourteen to sixteen are about the development and fluctuation of people within the age group of 15-24 years, composition of households and people migrating from one zip area to another. For the calculation of liveability the LM determines the zip code of a certain area and incorporates aggregated data regarding that zip code area in the calculation based on the indicators. The calculation depends on the presence and availability of data from the data sources within a specified area. The assumption is made that the attributed data is valid and could be generalized for the entirety of the Netherlands. On the basis of these estimates, a

---

[38] Image acquired from 'The Dutch Measure of Liveability: A Critical Data Studies Approach' by Smit and van Schie (forthcoming).

liveability score is calculated on a scale of a 6ppc (six number zip code) area with a minimum of forty inhabitants within that area (Leidelmeijer 2017, p. 9-10).

The indicators are constructed upon data which reflect level of income, unemployment, ethnicity, migration status, household details and other similar micro data. A certain normative perception can be extracted from these indicators, as there is strong focus on ethnic minorities and specified social groups which inherit a negative value for the calculation of the liveability within an area. This connotes that the social groups classified in one of the indicators which have a negative value for the liveability calculation are the problem groups. The use of the term "allochtones"[39] is a very loaded term and implies several problematic notions, as does the division based on ethnicity[40] and/or disabilities. The implications of these uses are less problematic in terms of the initial datasets, but more in how it represents a deeper cultural problem concerning a binary of 'civilized' versus 'non-civilized', 'healthy' versus 'non-healthy' and 'autochtoon' versus 'allochtoon' (Wekker 2016, p. 63). Consequently it is no surprise that neighbourhoods with a high level of 'allochtones' - such as Overvecht and Kanaleneiland - generally have a lower liveability score than neighbourhoods with a high level of 'autochtones', as visualized in fig. 7. Donna Haraway (1988, p. 577-578) adds to this with the argument that these technologies embed a binary opposition, in which *the norm* is compared with *the others*. The depicted *others* are "analyzed not as citizens, or even people, but as problems to be solved or confined" (Said 2014, p. 207). The construction of this binary opposition, and the knowledge produced by it, risks transforming into an alternation of *epistemic violence*[41] (Spivak 1988).

This is understood as an example of *standpoint theory* (Harding 2015). Within the LM that what is left *unseen* or what *standpoint* is used can be described as the normative perception of the "native working healthy Dutch citizen", which i.e. is inscribed in the LM by the exclusion of indicators concerning the native Dutch citizens, while focusing on various *other* social groups like ethnic minorities or the disabled which inherit a negative value for the liveability calculation (Smit and van

---

[39] On November first, 2016 the CBS and the Dutch Scientific Council for Government Policy stated to officially not use the term 'allochtoon' anymore within their records, documents and systems, as it is no longer sufficient to account for the diversity of the Dutch immigrant population (Bovens et al. 2016). It was also announced that there is an intention to gradually phase out the terms 'western' and 'non-western'. Still these terms are used in the four web applications, and it is now known if and when this will be altered.

[40] For a broader elaboration on the use of the term "allochtones", its origin, and the division used in the LM based on ethnicity see 'The Dutch Measure of Liveability: A Critical Data Studies Approach' by Smit and van Schie (forthcoming).

[41] Epistemic violence is a concept regarding a subject, object and action in which epistemological violence is attributed (Spivak 1988). Within this concept the subject of violence is the researcher and/or institutional actor/agent, the object is the other, and the action is the interpretation of data which is presented as knowledge (Teo 2010).

Schie forthcoming). This shows how the configuration of data and its indicators within these applications can implicitly construct a normative measure upon which socio-cultural groups are perceived as *others*, not complying the set socio-cultural norms and which are understood as the *norm*.

An example of appropriation of the data through configurations within the LM can be seen in fig. 10. Here the scores of liveability within the Netherlands are visualized, in relation with calculated scores in the first version of the LM.
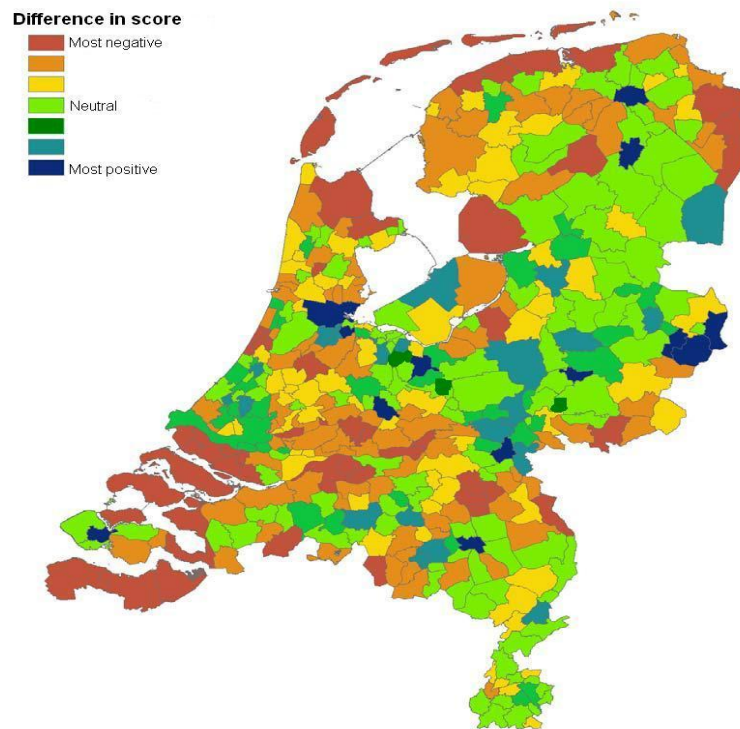
Figure 10: Differences in scoring of liveability between Leefbaarometer 1.0 and 2.0

The areas coloured red, orange and yellow mean that these areas are scored more negative in LM 2.0 in relation with the calculated scores of that area within the LM 1.0. The colours dark green, light blue and dark blue depict that the areas are scored more positive in LM 2.0 in relation with LM 1.0. This change occurred as the values for the five dimensions and the configurations of indicators were altered. The most notable change resides within the population and facilities dimensions. Within the LM 2.0 the population dimension has a value of 15% for the calculation of a liveability score, and the facilities dimension has a value of 25%. Within LM 1.0 the population dimensions had a value of 26%, and the facilities dimensions 17%. This means that the indicators within the population dimensions affected the liveability score much more than is the case in LM 2.0. Foremost rural areas as Zeeland and parts of Groningen, Friesland and Noord-Brabant are affected in a negative manner with this change in value of the dimensions, whereas most urban areas are affected in a positive manner.

Urban areas in general have more and better facilities in comparison with rural areas, while the population is more diverse in urban areas (CBS 2017). This could be one of the causes why rural areas are affected in a negative manner with the change in value of the dimensions. These variations show how changing relatively small variables within the system can affect the potential outcomes in a severe way, while this is not explicitly mentioned throughout the application and how multiple representations of the datasets can be constructed based on the given value to various indicators and dimensions.

4.1.2 The Geluksmeter

The GM application uses statistical data from the CBS containing statistical information on people's age, income levels, state of welfare, surroundings and the amount of social benefits in an area, originating from surveys held among the Dutch population regarding questions which examine subjects as happiness, satisfaction, and/or wellbeing (translated by author) (CBS 2016, p. 8-11). These are categorized, placed in classification scheme's and consequently translated into indicators regarding states and forms of happiness within the Netherlands. National averages are shown when the user enters the application and specifies variables as location, gender, education and age[42]. For example the question "hoe zit dat in jouw provincie?" is asked, which means that the user can see what the average happiness score is for his/her province. Clicking on this question is the only option the user gets to move forward within the interface. Next, a map of the Netherlands is visualized, on which the user can click to select a province to see its average happiness score (see fig. 11).
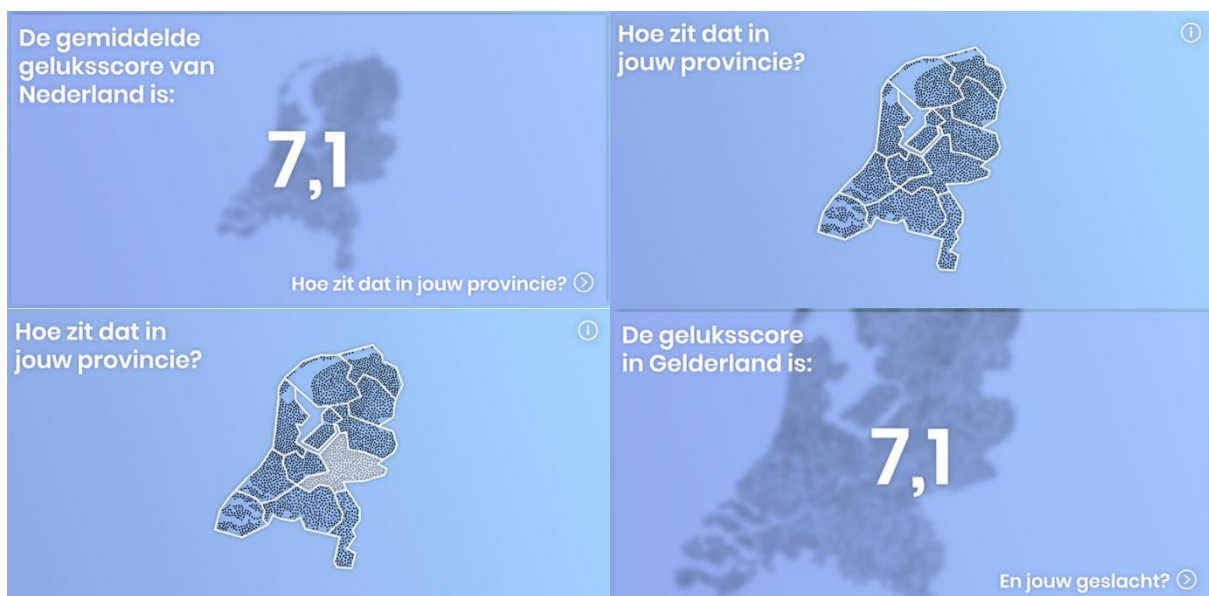
Figure 11: Geluksmeter start page and province selection

---

[42] See appendix 2.1 on page 59 for the full discursive interface analysis of the GM application.

From left to right, from top to bottom: (1) The start page of the GM interaction saying that "The average happiness score in the Netherlands is 7.1. What is it like in your province?" (translated by author). After clicking one is prompted with the (2) province page which invites the user to select their province by clicking on it (3), after clicking a province, the GM zooms in and displays the average happiness score of the region (4).

Next, the user has to specify their gender to progress (see fig. 12). A binary option is provided whereas the user can choose between male or female. Here, similarly as above, the interface forces you to choose if you are a male or female - further exemplified by the 'help' window, which emphasizes the binary option. You do not have the freedom to provide any other answers besides those which are provided by the interface.

Figure 12: Geluksmeter gender selection

From left to right, from top to bottom: (1) The user is prompted to answer "What is your gender?" (translated by author). If in doubt, one can click the information button in the top right corner, which unfolds a window with instruction (2) saying "click either the male or female icon" (translated by author). Through the hover animation (3), one is invited to click one of the options, after which the average happiness score of either Dutch males or females is displayed (4).

When a gender is chosen, the user has to further specify their personalized happiness score by selecting an age group in which they are situated. The age groups consist of: 18 – 24, 25 – 44, 45 – 64, and 65+ (see fig. 13). It is not stated why an age group below 18 years of age is not incorporated within the application. The only other interactive function is the "i" button on the top right corner of the screen. Clicking on this button shows a text which elaborates what the user has to do to continue in the system.

Figure 13. Geluksmeter age selection

From left to right, from top to bottom: (1) The user is prompted to answer "What is your age category?" (translated by author). If in doubt, one can click the information button in the top right corner, which unfolds a window with instruction (2) saying "click the hourglass representing your age group" (translated by author). Through the hover animation (3), one is invited to click one of the options, after which the average happiness score of either of the specified gender in that age category is displayed (4).

Through the affordances of the system it is understood that the user of the GM application is imagined as a Dutch resident, who is cisgender, and of age. These are assumptions the system makes about its users. However, even implicitly, there are some hidden norms embedded within the system as well. People of i.e. sixty-five years of age are put into the same demographic category as people who are eighty years of age and are long retired and live a less active life. Not all people above sixty-five years of age have the same lifestyle, which seems to be the assumption in the system - even though that lifestyle has quite the impact on how one experiences their lives.

Next, the user has to further specify their personal characteristics by selecting their level of education (see fig. 14). One can choose from: basis, VMBO, HAVO, VWO, MBO, HBO, and WO[43]. Presumably, the system refers to the 'highest' education level, but this is not specified as such.

---

[43] Possible choices are: primary school, high school (from lowest to highest level: VMBO, HAVO, VWO) and subsequently vocational education (MBO) or university of applied sciences (HBO) or university (WO).

Figure 14: Geluksmeter education selection

From left to right, from top to bottom: (1) The user is prompted to answer "What is your level of education?" (translated by author). Through the hover animation (2), one is invited to click one of the options, after which the average happiness score of one's gender, one's age group, and one's educational level is displayed (3). After this selection an average is given based on the user's personalization variables, and the GM first displays the highest and lowest values for your demographic, after which it shows the others as well (see fig. 15). All of the eight dimensions are visualized within a circular graph. If you use your mouse to hover above the visualization, different scores are presented based on the eight dimensions.
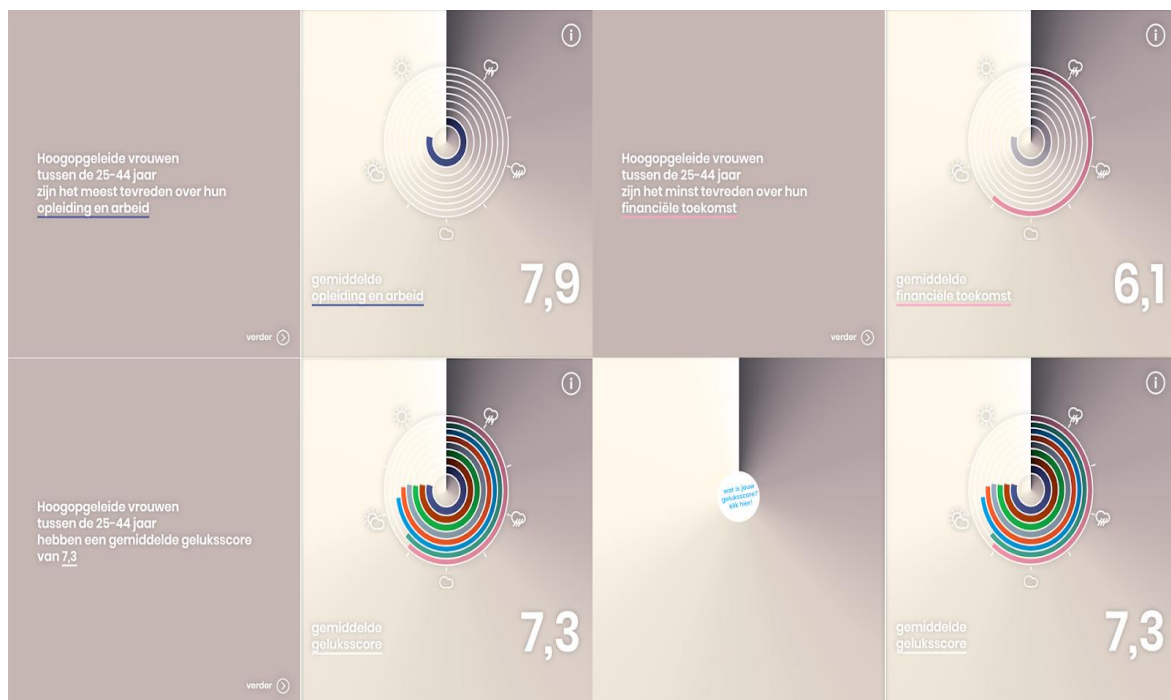
Figure 15: Geluksmeter average scores for the chosen demographic

From left to right, from top to bottom: (1) The user is presented with the theme they are most satisfied about, after which (2) they are presented with the theme they are the least satisfied with. Subsequently, (3) all of the measures are displayed. Finally, (4) the user is invited to fill out the questions themselves in order to calculate a personal score. No further information is given how these dimensions are calculated or visualized. The user is not given an option to gain deeper insights into the calculation or representation of the used data. The interface is thus verily simplistic in its functions, as the user only has a few interactive options in each screen which it can alter and use to navigate through the system.

4.1.3 The Uitkeringsmonitor

The UM application uses statistical data from the CBS and CWI, enriched with data from the GBA. The CBS counts the population receiving social welfare/benefits within a specified area, turns this into measures, constructs baseline rates, develops interrelations between factors concerning social welfare/benefits recipients, and consequently transforms this into graphical means within the Venn diagram and the geographical map of the Netherlands (see fig. 6). In appendix 2.3 on page 78 a complete walkthrough of the UM interface is shown as described in the 'Walkthrough Method' (Light et al. 2016). The categorization of people within six categories of social welfare/benefits recipients is regarded as a form of *normalization*[44]*,* as the people are assessed, processed and grouped according to certain norms wherein profiles of people are constructed, upon which they are sorted with specified characteristics and classifications into six categories of social welfare/benefits recipients, and their geographical characteristics. As it concerns sensitive data, it only visualizes the amount of social welfare/benefits recipients on a municipal scale. The user is allowed to compare the data from three categories of social welfare/benefits recipients within a Venn diagram (see fig. 6). A timeframe of a specific month has to be selected, together with a year ranging from 2010 to 2015. The user can download the representation in a .pdf format, but by doing so the visualization loses its interactive features and becomes static.

4.1.4 The BenIkGemiddeld

The BIG application does not have official documentation regarding the use of data or reflection on its functioning, as a consequence it is not known what exact data sources the application uses. The application does forward you towards general reports of the CBS in which national figures are described on various societal numbers regarding averages generated by surveys held among the Dutch population with sample groups, simulations and regression analyses. The user has to answer questions regarding five dimensions: culture and leisure, population, expenditures, health, and

---

[44] The term normalization resides within the concept of "engines of discoverability" as posed by Hacking (2007), and seeks to fashion the world to fit certain norms (Kitchin and Lauriault 2014).

income (translated by author). When the answers are given and a specific demographic group is selected based on gender, age and income, the user gets to see visualizations in the form of a chart, bar or a weight scale on which is illustrated if the user belongs to an average demographics group or not (see fig. 16).
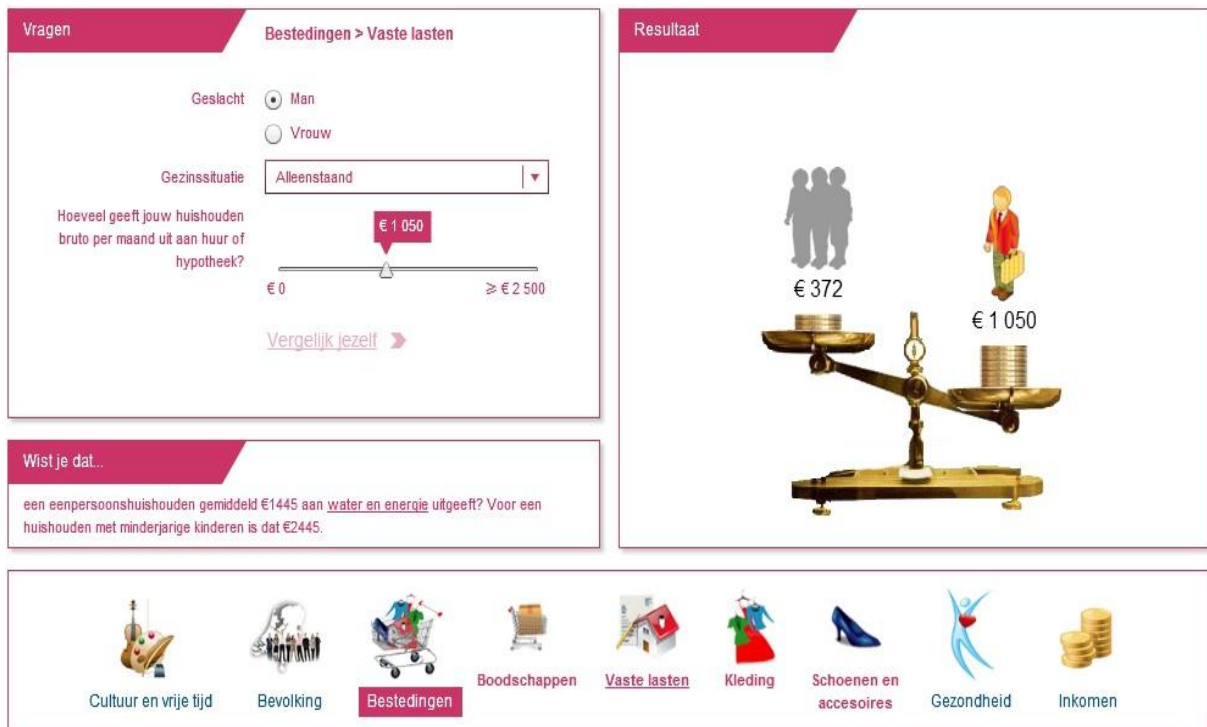
Figure 16: Comparing the user with the national average

Depending on the chosen theme, an additional question has to be answered, as seen in fig. 10 the theme "expenses" (translated by author) is chosen. Here an additional question regarding your family status has to be answered, together with a bar on which the user has to answer how much its household spends each month on rent or loans. After answering the user can compare itself to the national average regarding the provided answers. Furthermore, a piece of information is given regarding the selected theme. It is unknown how the personal calculations are constructed, or how the national averages are provided.

The application enforces the user to pick a persona visualized as a small cartoon figure, in which the occupation of this persona is depicted using a small instrument like a paintbrush, pen, suitcase, hammer and such (see fig. 5). This illustrates which occupations and kinds of people are included in the system and thus have a bigger chance to belong to the explicitly constructed *norm* visualized by the application. Furthermore, only the choice of gender between male/female is given, and no persona can be selected without a distinct job or education. No contextual information is given how this average numbers are calculated, or how the scores attributed to the variables given by the user are constructed. The knowledge produced through this web application shows in an

explicit manner what the societal *norm* regarding the five embedded theme's is, and if the user conforms to this normative measure. Within this application the norm is made very explicit, and the user can even interact and experiment with the set norm. But, as the user has to choose a set persona to continue within the interface, the norm is forced upon the user as it does not have the freedom to develop and customize its own persona which resembles the personal characteristics of the individual[45].

4.2 Graphic design

The graphic design of the four applications is examined following Kennedy's (2018) five key points to look for in visualizations. The following schematic table is constructed upon these five features and described regarding the graphical designs of the applications (see fig. 17), whereas examples of the applications and their graphic designs are examined.

Applications	Data Represented in/as	Data Source Available/Known	Key Legend available	Axes & Scales Ranges and values	Title Presented
Leefbaarometer	Geographical maps Bar charts	CBS / GBA In documentation Link to help pages	Legend depicting nine colours as scores	Four geographical scales Time scale Axes only in bar charts	Presented in legend and on top of the geographical map
Geluksmeter	Circular bar charts Circular graphs	CBS In documentation	Legend depicting eight colours as dimensions	No axes Circular scale ranging from 1 to 10	Only at beginning and end results
Uitkeringsmonitor	Geographical maps Venn diagram	CBS / GBA In documentation	Within Venn diagram Beneath map	One geographical scale Time scale No axes	Presented above the map and in Venn diagram
Ben Ik Gemiddeld	Graphs Weight scales Bar charts	Not available	Not available	Time scale Weight scale comparing financial numbers Axes in graphs and bar charts depicting trends	Multiple titles regarding the content of each window

Figure 17: Overview of graphic features

---

[45] For the full discursive interface analysis of the BIG application as described in the 'Walkthrough Method' (Light et al. 2016) see appendix 2.4 on page 83.

4.2.1 The Leefbaarometer

The vision within the LM is spatially affected according to the scale on which the user sees the map, as a large part of a postal code area can be visualized with a red colour, while when zooming in on grid-level only one or two neighbourhoods are actually depicted with the red colour. This is noticed within fig. 29 - 32 in appendix 2.2.1 on pages 66 - 69. This suggests that the red scored neighbourhoods are prioritized over the green scored neighbourhoods, which further exemplifies that the system has a particular focus on societal elements which are attributed to a devaluation of the liveability measure. This could affect the epistemic process, as the visualized scores are not one-on-one translations from the initial data. They are altered and configured in comparison with a non-transparent constructed national average, which only becomes clear after clicking on a scored area and reading the texts. It is thus implied that all scores are regarded within the perspective of the norm, and entails if citizens living within an area do not conform to this norm, they are defined as 'others' and scored in a negative manner regarding the liveability within that area. This is understood as appropriations of the used data as the scoring mechanism is built with a strong bias towards the societal norm.

4.2.2 The Geluksmeter

The GM embeds this bias in a similar manner, but different shape. The definition of a construct like happiness and what is acknowledged as happiness is transformed into simplistic circular bar charts, in which each dimension attributed to happiness has a specific colour. The more the user complies to the represented conception of happiness, the higher the score and the fuller coloured the circular bar charts become. The circular bar charts gaining more colour as the user answers the stated questions with the 'right' answers, connote that the system has a reference frame in which the answers are related to. If the answers fit within this frame, the user gets a higher score. If the answers do not fit within the frame, the user gets a lower score. The user is forced to answer each question on a spectrum from one to ten, on seemingly simplistic questions regarding very complex subjects as your potential future financial situation. No context is given on these questions and no alternative answers are provided.

4.2.3 The Uitkeringsmonitor

Regarding the graphic design of the UM the user can alter the visualized forms of allowances with a pallet of sixty colours, and consequently notice which colours appear on the geographical map relating to the Venn diagram. The user has to select one of seven choices which is then attributed to one of the three circles in the Venn diagram (see fig. 18). The colour which is attributed to a specific dimension and is shown on the map is the form of social benefit/welfare which the most people in that area receive. The estimate is given of a measure per one-thousand inhabitants within a

municipality. What is problematic within this graphical design is that the Venn diagram only incorporates three different dimensions of social welfare/benefits, while there are six in total. The user can thus never visualize a map with all allowances in separate dimensions. The Venn diagram can thus in theory be regarded as a useful visualization tool, but when put in practice with a dataset which is classified and categorized in more than three dimensions, it lacks the possibility to represent more than these three different dimensions.
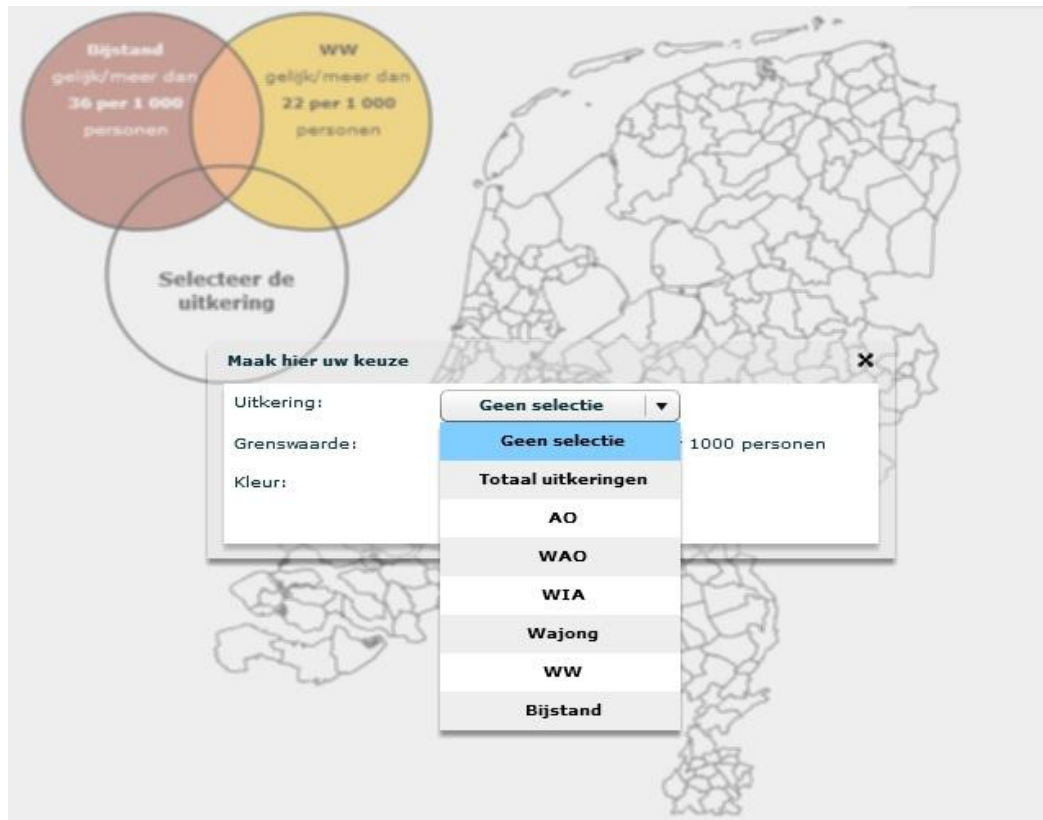
Figure 18: Variables Uitkeringsmonitor

Furthermore, the data is represented on a geographical map of the Netherlands, with no contextual information, axes or scales except a possibility to select a time frame of a month and year and a selection of six social welfare/benefit types (see fig. 18). The data source is not made explicit throughout the application, as the user is forwarded towards official documentation of the CBS where the data source can be noticed. The key is presented within the Venn diagram, as it depicts a measure of social welfare/benefits recipients per thousand inhabitants and which colour is attributed to that group. Additionally a legend is presented below the map concerning national figures on the population of the Netherlands, and when the user hovers its cursor above a municipality information on the working population and numbers on social benefits/welfare recipients within that area are shown on a scale of one per thousand inhabitants (see fig. 19). This is the only contextual information the user gets to see within the interface. One primary title is available above the

geographical map stating "persons with an allowance per municipality" (translated by author).
Secondary titles are presented within the menu with regard to forms of allowances and location.
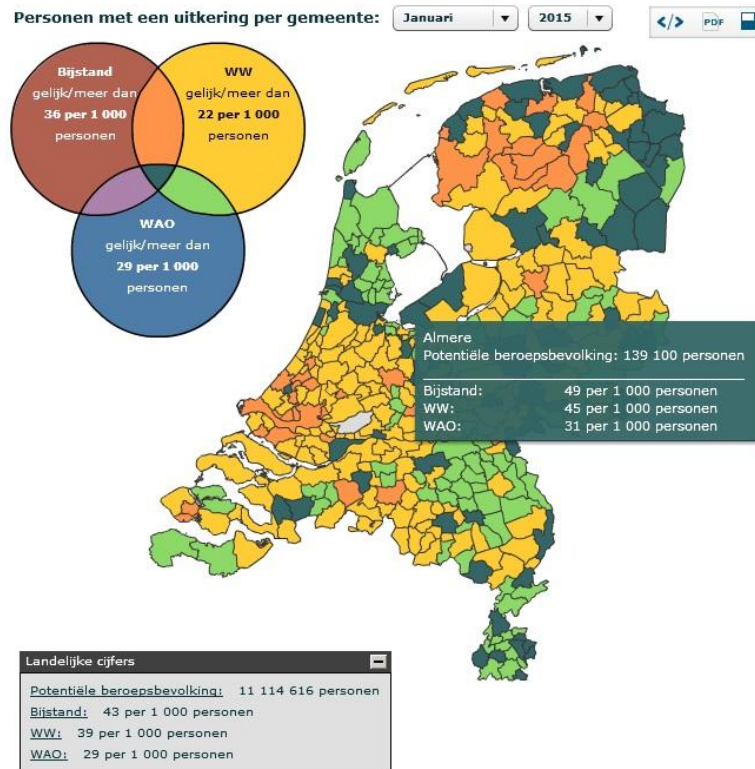
Figure 19: Uitkeringsmonitor Map Additional Information

4.2.4 The Benikgemiddeld

The BIG makes the norm explicit within the visualizations and the interface itself as it starts with the
question "how average or unique are you?" (translated by author) as shown in fig. 16. This connotes
that if the user is situated within the national average of a given theme, it is categorized as
"average", while the user that is positioned away from the national average is considered "unique".
The term "unique" may sound like a positive alternation of the *others*, but becomes clear when
focusing on the visualization which is constructed when a user answers a question and is categorized
as "unique". A weight scale is visualized with two values, on the left the "average" class is depicted,
while on the right the "unique" class is visualized. As the "unique" class always is a minority in
comparison with the national average, this group is always outbalanced in the visualization by the
norm. The visualization illustrates that the national average outweighs the *others,* and visualizes it as
the norm is depicted above the *others.* This very literally connotes that the norm is prioritized above
the *others,* and that they are higher on the metaphorical social scale than those not fitting within the
norm.

The affordances and graphical designs of the four web applications are somewhat similar in

their interactive visualization means and use of maps and charts. The GM and BIG resemble to each other in visualization elements as visualizing data in bars and charts, as do the LM and UM on geographical maps. Where the BIG uses a wider array of visualizations embedding graphs, bar charts and scales, it is also explicit about its functions and its inscribed norms, while the GM, LM and UM are more subtle and function more as a black box. The UM works with sensitive data on social benefit/welfare recipients, singling out these people and categorizing them in terms of allowances. The LM functions in a similar manner, but focuses on social groups which potentially devalue the liveability measure of a certain area. The GM defines the construct of happiness according to eight dimensions, and instigates to show the user a 'personalized' score by marginalizing the opaque concept of happiness into eight questions. Within the applications norms are made explicit with the use of average and national figures/numbers, which are visualized in charts, bars graphs and/or maps. But these norms are also implicitly embedded within the systems i.e. in the form of the categorizations, indicators, dimensions and variables used to construct the visualizations and its vision.

4.3 Environment of expected use

In the analysis of the environment of expected use I follow the same structure as noted by Light et al. (2016). The first part focuses on examining the vision of an application. Within this vision the purpose, target user base, and scenarios of use are points of inquiry. These are mostly communicated through the technology's organisational materials and its interface (Light et al. 2016). The second part is the technical walkthrough and is the central data-gathering procedure. The technical walkthrough of the interfaces are depicted in appendix 2.0 on pages 58 – 88.

4.3.1 Vision Leefbaarometer

The documents 'Leefbaarometer' (CBS 2017); 'Leefbaarheid in balans' (Leidelmeijer et al. 2011) and 'RIVM Rapport 630950002 Kwaliteit van de Leefomgeving En Leefbaarheid' (Leidelmeijer et al. 2008) regard motivations, reflections and configurations of the LM. When analyzing the reports, focusing on the purpose of the LM, it is articulated that the report wants to inform the Dutch public about the measure of liveability based on hundred indicators within five dimensions. Moreover, it instigates to make the research activities of the CBS transparent and show the results of their research with according held surveys among the Dutch public. The primary purpose of the report is to map the status of liveability within the Netherlands and reflect on it (CBS 2016). Within these reports a definition is shown of how liveability is defined based on several literature studies (Leidelmeijer and Kamp 2008; Leidelmeijer et al. 2017). These are at the base of the liveability configuration, conducted in years before the construction of the LM. Based upon these studies, the following definition of liveability is used: "the liveability is the measure in which the surrounding connects to

the wishes and demands set by people in relation to a liveable environment" (Leidelmeijer et al. 2017, p. 11; translated by author). This is paradoxical as the LM is framed as an application which shows an objective representation of liveability within the Netherlands: "The LM provides an objective judgement on liveability, constructed upon the values regarding liveability given by inhabitants" (translated by author) (Leidelmeijer et al. 2017, p. 9). Not only is it paradoxical, it gives a skewed perception of liveability as within the interface of the LM only a description is shown that the liveability is constructed upon hundred indicators, and not what the system defines as "liveable". The function of the LM differs with the other applications in so much that it is not solely designed to inform the public, but also to inform policy and possibly to be used by governmental actors. In an interview[46] with one of the designers of the LM it was made clear that the application was primarily created to be used by municipalities within the Netherlands (Smit and van Schie forthcoming). This is an entirely different public than the general public, with a different level of expertise, knowledge, data literacy and experience concerning statistical data. Furthermore, this intended audience is not made explicit throughout the system, as in the 'RIVM Rapport 630950002 Kwaliteit van de Leefomgeving En Leefbaarheid' (Leidelmeijer et al. 2008) it is stated that the main purpose of the LM is to give the general public insights into complex statistical data. This specific target audience cannot be found in any official document regarding the LM. The target user base could thus be described as Dutch civilians who would like to inform themselves on the status of liveability within the Netherlands and would like to reflect their personal score to the average scores in the Netherlands. Additionally, as described in the held interview, the constructors of the application have an imagined audience in mind which consists of municipal employees, and have a very different level of data literacy and level of interpretation of the visualized data. There are no data known on the usage of the LM by the public, and what the demographics of the user base are. It begs to question if the other applications are constructed with this same imagined audience in mind, as they differ in functionalities while embedding similar visualization and design practices.

When focusing on the translation of this report in the LM, it can be noticed that only a few elements are translated literarily. I.e. the dimensions with according indicators are all similar, as are the calculation methods. The visualization practices are somewhat different. In the report only bar charts and graphs are used to visualize the statistical data on the liveability, while in the LM application all the data is visualized on a geographical map. The report 'Leefbaarometer' (CBS 2017) elaborates that this is done to make the data interactive and better apprehensible because the broader public can immediately see on which area of the Netherlands the data relates to. It is not known how the data are divided on the geographical map, and what variables and measures are used

---

[46] See appendix 4 on page 86 for a summary of the interview (in Dutch).

to place specific datasets and data points on particular positions on the map. This mystifies the translation process of the initial data into graphical form.

The purpose and target user base have been described, now the last element within the vision of the interface remains, which consists of the scenario of use. The theoretical scenario of use beholds users whom would like to see the average level of liveability in the Netherlands, and find out what their personal score would be in their environment. The user can personalize his/her score based on criteria as gender, age and location. Additionally, various averages are given based on the five dimensions of the 'Leefbaarometer' (2017) report. The practical scenario of use within the technical walkthrough can be seen in appendix 2.2 on pages 63 – 78. The governance of the LM, which consists of how the LM seeks to manage and regulate user activity in order to sustain its operating model and fulfil its vision is shown in appendix 2.2.3 on page 76. The operational model and technical features of the LM application are shown in appendix 2.2.4 and 2.2.5 on page 78.

4.3.2 Vision Geluksmeter

The report 'Welzijn in Nederland' (CBS 2016) can be regarded as an official document describing the construction of the GM. When analyzing the report, focusing on the purpose of the GM, it is articulated that the report wants to inform the Dutch public about the measure of happiness, based on the eight dimensions: present financial situation, future financial situation, health, leisure, social life, government, safety, and your living area.. Moreover, it instigates to make the research activities of the CBS transparent and show the results of their research with according held surveys among the Dutch public. The primary purpose of the report is to map the state of wellbeing of the Dutch population (CBS 2016). When the focus of inquiry is shifted from the report 'Welzijn in Nederland', to the GM another purpose becomes visible. It wants to make the complex statistical data from the CBS research more easily accessible and readable to a wider public by visualizing it in interactive graphical formats. It tries to do this by visualizing the data on the map of the Netherlands with simplistic circular graphs with according scores based on the eight dimensions of the report. The statistical data are thus translated in graphical form to make it comprehensible for the public.

The document 'Welzijn in Nederland' (CBS 2016) elaborates on the motivations and definitions used within the GM, and elaborates on the held surveys on the measure of happiness within the Netherlands. What is noticeable is that the document defines different dimensions for the measure of happiness than the GM web application in digital form. This means that different indicators, and perhaps even data categories are used to calculate the measure of happiness. Furthermore, the document does not speak of a happiness score (geluksscore), but of an index on wellbeing (welzijnsindex) (CBS 2016, p. 9-10). The CBS contextualizes the term wellbeing as a subjective term with three aspects. The first one is the aspect of evaluation, which is measured by

the degree of satisfaction with (parts of) life. The second aspect consists of the affects of people, which is defined as positive and negative emotions towards life. And the third aspect consists of the eudemonic aspect, which means how valuable certain things in life are to people. According to the CBS these three aspects define the degree of wellbeing in the Netherlands, and the primary purpose is to map and show this divided into the eight dimensions. No further motivation is given why these three aspects were chosen to define the wellbeing of the Dutch population. Furthermore, it is not clear how these three aspects precisely relate to the eight dimensions on which the GM has been built. This mystifies the precise definition of the happiness score as used in the GM web application, and constructs the term as a vague and opaque conceptualization of the wellbeing term. The CBS does acknowledge that the application cannot be regarded as an objective system simply reproducing the construct of happiness according to eight dimensions (CBS 2016). The CBS is conscious that such a complex construct cannot be defined and calculated by eight simple questions, as it is stated that the main purpose of the application is to inform the public on the state of happiness within the Netherlands, and let the user experiment with different dimensions within the concept of happiness (CBS 2016, p. 9-11).

Now that the purpose of the application has been established, the analysis shifts towards the target user base of the GM. The target user base could be described as Dutch civilians who would like to inform themselves on the status of happiness within the Netherlands, are interested in their personal happiness score, and would like to reflect their personal score to the average scores in the Netherlands. There are no data known on the usage of the GM by the public, and what the demographics of the user base are. As the GM is running on statistical data gathered by the CBS in 2015, and no update have been made since, it is logical to think that the more time progresses, the older the data gets, and the value of the scores will decrease as it will less likely represent the current state of affairs within the Netherlands. This adds to the problematic notion of the GM, as it does not clearly specify that the calculated scores are based on data from 2015. Users could thus interpret the scores as a present representation of happiness, and give them a skewed perspective on the actual measure of happiness.

The purpose and target user base of the GM have been described, now the last element within the vision remains, which consists of the scenario of use. The theoretical scenario of use beholds users whom would like to see the average level of happiness in the Netherlands, and find out what their personal happiness score would be. Additionally, various averages are given based on the eight dimensions of the 'Welzijn in Nederland' (CBS 2016) report. No motivation is given how these scores are calculated or how the dimensions are defined, and why. The practical scenario of use within the technical walkthrough can be seen in appendix 2.1 on pages 59 – 62. The governance of the GM, which consists of how the LM seeks to manage and regulate user activity in order to

sustain its operating model and fulfil its vision is shown in appendix 2.1.3 on page 62. The operational model and technical features of the GM application are shown in appendix 2.1.4 and 2.1.5 on page 63.

4.3.3 Vision Uitkeringsmonitor

The report 'Arbeidsmarkt in Cijfers' (CBS 2016a) can be regarded as official documentation regarding the UM. This report is not explicitly mentioned by the application as documentation regarding the application, but can be found when the user is forwarded towards the open data Statline platform of the CBS. When analyzing the report, focusing on the purpose of the UM, it is mentioned that the report wants to inform the Dutch public about the state of the working population, social welfare and numbers on income levels and work rates. Moreover, it instigates to make the research activities of the CBS transparent and show the results of their research with according numbers. The primary purpose of the report is to map the state of the working class and income levels of the Dutch population and additionally provide a picture of social welfare and benefits recipients (CBS 2016). The CBS and the Cultural Planning Bureau (CPB) cooperatively examined the status of social benefits/welfare recipients within the Netherlands, on which the UM has been constructed. The division in the six categories of social welfare/benefits recipients within the UM is based on official labels of these various welfare/benefits accounts. These are matched with demographics data from the CBS and GBA to localize where which recipient lives. As the document 'Arbeidsmarkt in Cijfers' (CBS 2016a) notes that this kind of information is sensitive and should be handled with care, it is interesting that still such a web application as the UM is constructed upon this data. It begs to question if the people represented through this application gave consent to this form of representation. When the focus of inquiry is shifted from the report to the UM application, another purpose becomes visible. It wants to make the complex statistical data from the CBS more easily accessible and readable to a wider public by visualizing it in interactive graphical format. It tries to do this by visualizing the data on the map of the Netherlands divided into different categorizations of social welfare/benefits recipients.

Now that the purpose of the application has been established, the analysis shifts towards the target user base. The target user base could be described as Dutch civilians who would like to inform themselves on the status of social welfare/benefits recipients within the Netherlands, and are interested in where specific social welfare/benefits recipients live within the Netherlands. No data is known on the usage of the UM by the public, and what the demographics of the user base are. As the UM is running on statistical data gathered by the CBS in 2014 and 2015, and no updates have been made since, it is logical to think that the more time progresses, the older the data gets, and the value of the application will decrease as it will less likely represent the current state of affairs within the

Netherlands. Users could thus interpret the visualizations as a present representation of social welfare/benefits status.

The purpose and target user base have been described, now the last element within the vision of the interface remains, which consists of the scenario of use. The theoretical scenario of use beholds users whom would like to see the what social welfare/benefits recipients are most present in the Netherlands, and experiment with different types of social welfare/benefits recipients and locations. The user can personalize parameters based on type of social welfare, colour, measure and location. No motivation is given how this data is visualized or how it is calculated. A small text next to the visualization relates to the visualized data based on set parameters. The practical scenario of use within the technical walkthrough is shown in appendix 2.3 on pages 78 – 71. The governance of the UM, which consists of how the UM seeks to manage and regulate user activity in order to sustain its operating model and fulfil its vision is shown in appendix 2.3.3 on page 81. The operational model and technical features of the UM application are shown in appendix 2.3.4 and 2.3.5 on page 82.

4.3.4 Vision Benikgemiddeld

The BIG application does not explicitly refer to an official document regarding the definitions, motivations, choices and framing of the application. The implicit reference is made by forwarding the user to the annual demographic reports of the CBS (CBS 2017a), and the Open Data Statline platform. As such, the functionalities and motivations of the inner workings can be regarded as a black box. The framing of the application can be interpreted in a similar manner as the Open Data Statline platform, as the reference is made that the application is an extension of this platform. As the Open Data Statline platform is framed as a *neutral* system to gain direct access to open data provided by the CBS, the BIG application should be understood in the same manner (CBS 2017). As the BIG application is not accompanied with official documents and/or reports, no organisational materials can be analyzed to establish the vision of the application. As such, the vision has to be established solely through the scenario of use and technical walkthrough. A hypothetical vision can be extracted from this walkthrough whereas users can experiment with set norms based on themes such as finance, health, leisure, education, safety and governance. The practical scenario of use within the technical walkthrough is shown in appendix 2.4 on pages 83 – 87. The governance of the BIG, which consists of how the BIG application seeks to manage and regulate user activity in order to sustain its operating model and fulfil its vision is shown in appendix 2.4.3 on page 87. The operational model and technical features of the BIG application are shown in appendix 2.4.4 and 2.4.5 on page 88.

4.4 Epistemic notions and critical considerations

The GM, LM, BIG and UM call forth data publics which are incited to practice their own analyses of a vast amount of quantitative data from the CBS, through social, material and literary (especially visual)

means. The applications generate users who can enact their own experiments, appropriate these and share their results and produced knowledge to others. The data publics are affected by the socio-technical arrangements in which they are embedded, configured by complex flexible arrangements of various actors which are mobilized and enacted by the available and used open data (Ruppert 2015, p. 130). It is through the actions and interrelations of elements within *the looping effect* that insights into the data is given (Hacking 2007). The applications *make up* people with according reference frames through which the state and its data is enacted. These processes only partially make the state transparent, but foremost produce a *meta-social reality* in which only prioritized aspects of societal phenomena are represented. This is especially true within the LM and GM applications, as drawing from Harding's (2015) theory a clear *standpoint* within the design and representation of data is noticed. This standpoint concerns the 'native healthy working Dutch citizen'. The applications draw boundaries and restrictions around what is included and excluded in this constructed reality, displace implicit epistemic notions and reconceptualise institutional recording practices within an *objective* and *neutral* frame. The key issue is that the imagined audiences of these technologies are not always data literate people, or peers residing within institutional spheres apprehending the complex configurations.

Regarding the complexity of the systems, it can be argued that the GM and UM are constructed with an intended audience in mind which entail a lower level of data literacy, as these systems have fewer interactive options, personalization variables and the graphic designs are less complex as the user cannot click on the visualizations to gain more information or see other bars/charts regarding the visualized data. This means that the visualizations have to be taken as is, whereas in the LM and BIG the user is forwarded towards more information regarding the visualizations and has more options to experiment and play with variables and parameters.

Rather than only visualizing and representing societal phenomena, the web applications enable publics to see and analyse the data right in front of their eyes, as well as interact with, adjust, manipulate and experience it. These web applications thus bring dispersed phenomena such as invisible statistical numbers to publics and render it interactionally present. Therefore, it is through these *data assemblages* (Kitchin and Lauriault 2014, p. 5) that publics are not only produced, but also that the state is made interactionally present and experienced. The four web applications could thus be regarded as vehicles, in which the political transforms into the *neutral*, being technically embedded within the visualization technologies, which bears a strong relation to the concept of *governmentality* (Foucault 1991). While these systems differ in their functionalities and visualization practices, they all are situated within, and simultaneously echo neoliberal forms of g*overnmentality* (ibid.). Such data-driven technologies interweave quantification and computing into the realm of opaque socio-cultural norms and values and instigate to reconceptualise and represent these norms

and values through interactive visualizations. By doing so, they alter *governmentality* from disciplinary forms to instances of social control (Kitchin et al. 2017, p. 3-4).

Two knowledge systems are at work here: (institutional) techno-science knowledge on the one hand; and rendered fluid knowledge on the other. How should these knowledge asymmetries be understood? They could be conceived as expressions of *power-knowledge,* wherein certain epistemological preferences and reference frames dominate others. It is assumed that institutionally generated and visualized data assures objectivity, while the contingent, ambivalent and context dependent assemblage of practices embedded within this generation, transformation and interpretation of data is only partially apparent in the examined four *data assemblages* (Kitchin and Lauriault 2014, p. 4-6). Thus, "seeing like the state" shifts towards "seeing the state" (Ruppert 2015, p. 128), in which the rendered sight is affected and constructed upon the appropriation of the initial data and embedded socio-cultural norms and normative assumptions. Haraway (1988) defines this as the *god trick*, in which the persistence and objectivity of vision inscribed in technologies is questioned as "seeing everything from nowhere ... promising vision from everywhere and nowhere equally and fully" (Haraway 1988, p. 581-582). The applications are thus regarded as somewhat helpful tools to gain insights into complex statistical data, but embed problematic implications residing within their graphic design and normative conceptions of social realities and its values. Haraway's (1988) suggestion to hold designers accountable for the visualization practices and politics which are inscribed in the systems of vision, especially when the designers are governmental actors, is an argument which must be considered.

5. Discussion

This research shows how the use of data assemblages as a methodological disentanglement of data-driven technologies enables to make the multi-layered dimensions of these systems visible. As shown with the four web application, the transference of complex statistical data into graphical form has various implications regarding the applications as knowledge production entities, especially when the data is appropriated by third-party applications as the Allochtoon-o-meter. Such applications change the context in which the initial data is collected and published, and frame it in an ideological manner as fearing 'allochtones'. The data publics experience simplistic translations of statistical data within graphic designs, which are actively constructed to smooth out the implicit contextual knowledge of the initial data. As a consequence the technology becomes tendentious. It matters in which context the open data is deployed, and on what platform the applications are constructed (i.e. Geenstijl). In comparison, the web applications on the ODS platform from the CBS have official documents regarding the construction of the indicators and legitimacy of the data sources. As such the CBS seeks

to provide as much transparency as possible. The problematic implications within the four analyzed applications reside less within the fact that they are black-boxed or opaque, and more within the case that they are positioned as neutral and non-biased systems which enable the production of 'objective' accounts of 'facts'. Regarding the LM, it claims to present an objective account of liveability within the Netherlands, which would mean that there is a causal relation between ethnic minorities, or people with a disability within an area, and a lower standard of liveability. It begs to question if this scenario touches upon the actual social reality. As I have shown in section 4.1.1, when the indicators within the population dimensions of the LM are scrutinized, it becomes apparent that normative assumptions are at the base of these constructed indicators and that these are everything but all-inclusive or non-biased. Investigating these embedded normative conceptions enable to make visible that what is 'behind' the data and what is excluded from the system. Strathern (2000) calls this the 'tyranny of transparency', as these applications seek to provide transparency regarding the state, while simultaneously eclipsing certain social realities and prioritizing others. Providing transparency is thus a paradoxical effort, as only rendered versions of the state are shown.

The four web applications cannot be regarded as neutral mathematical entities quantifying societal phenomena, but as versatile epistemic technologies using performance indicators, accounting, ranking, benchmarking, sorting and other New Public Management practices to measure, value, represent, appropriate and visualize open data (Merry and Rottenburg 2015, p. 6). Socio-cultural norms and values are imposed on the user, as the user does not have the freedom to decide which norms are embedded in the system, or how these norms are constructed within the design. Their visualizations embed a veneer of *neutrality* and *objectivity* with the use of specific visualization conventions. These concerns are in line with the field of CDS, which argues that such data-driven technologies make all information equal and conform it to their graphic designs, neglecting the legitimacy of the source, transparency concerning the transformation of statistical data into visual form, and the contextual value of the initial data (see Dalton and Thatcher 2014; Kitchin and Lauriault 2014; Iliadis and Russo 2016). As such, the primary focus should be held on how the used data is collected, processed and published by the CBS; the potential dangers of institutional categorizations and constructed indicators misleading the knowledge produced; and embedding hidden biases. The web applications are a great opportunity to give insights into complex statistical open data, while simultaneously misinforming the public through the authoritative frame of governmental bodies and its visualizations, whereas the data can be misinterpreted by the wider public.

Knowledge produced through these data assemblages is thus understood as *performative*, as societal realities are appropriated according to political, technological and socio-cultural frames (Kitchin and Lauriault 2014). The multiple layers of the data assemblage each render and affect the used data in various manners. The understanding regarding the epistemic asymmetries relating to

these multiple layers should be a key focus point in how data-driven technologies and their imagery of vision use and represent open data. A set of rules needs to be stated on how sound data practices can be established, and accountability for presented and translated data into graphical forms needs to be taken. Furthermore, an understanding of how data is read and interpreted by the imagined audiences needs to be developed in order to diminish the misuse and misappropriation of open data. The following points can be understood as results of the performed analyses and enable the awareness concerning the use and representation of open data:

- *The context of how the open data has been collected, documented, processed, translated and visualized has to be made transparent.* Only when this is incorporated in official documents regarding the data-driven technologies, or within the interface itself, the user can get a deeper insight into the workings of the system and how it uses open data, what the initial data means and how it is represented;

- *The seemingly self-evident categorizations, classifications, indicators, dimensions and visualizations have to be made transparent and justified.* The designers should clarify why certain terms and definitions are used for categories, indicators and dimensions, and not others. This would enable to make their constructed nature more transparent. In regard of the visualizations, i.e. the 'Seeing Data'[47] project developed by Helen Kennedy (2018) could be taken into account to scrutinize graphic design choices;

- *The reproduction and appropriation of knowledge by imagined audiences of data-driven technologies should be acknowledged and described*. Not all appropriation can be known beforehand, but by understanding which publics are most likely to use the data-driven technologies, an educated guess can be constructed in which a better understanding of the imagined audience is formed. This could result in the diminishing of misinterpretations of the knowledge produced through data-driven applications;

- *The embedded socio-cultural norms should be described in a manner in which it becomes clear how the data-driven applications prioritize certain socio-cultural norms, while neglecting others.* This could be done by stating a hierarchical structure in which the vision of the instrument regarding certain norms is depicted. Then it would be possible to make explicit why certain norms are prioritized over others, and the user could interpret the data while being aware of the hierarchy.

Sound data practices can be realised along the lines of these four points, while understanding data-driven technologies as data assemblages drawing from the emerging and increasingly relevant field

---

[47] For more information on the Seeing Data project visit https://www.seeingdata.org

of CDS. As such, this research shows how to operationalize the concept of the data assemblage and answers calls residing from the field CDS to unpack these data assemblages, and uncover their multi-layered arrangements and normative configurations (see Draude et al. 2017 p. 1-3; Iliadis and Russo 2016 p. 7-8; Kitchin et al. 2017, p. 18).

6. Conclusion

The affordances of the examined web applications show the issue of hidden bias in data publication and (re-)presentation. The tools provided by the CBS on their ODS platform are a great opportunity to give insights into complex statistical data, but are vulnerable to misinterpretation and misappropriation. The issue at hand is the manner how the technologies we understand as neutral actually manifest our biases and place it within a discourse of objectivity and neutrality. As people generally think of data as objective, it should not be forgotten that the used and presented data are always a partial sum of all available data and thus never all-inclusive or fully representative.

The examined four web applications afford the scoring and classification of citizens, representing measures of social values like happiness or liveability, informing on the state of social security, and presenting the societal norm within interactive graphic designs. The provided open data can potentially be appropriated for ideological means, both intended – as in the Allochtoon-o-meter - and unintended – as in the LM – through its algorithmic bias and constructed non-neutral indicators. The unintentional side effects facilitate a skewed representation of society, while claiming to merely give objective accounts of reality. The insights gained through the four applications seem to depict social realities in which the socio-cultural norms dictate what is seen as positive or negative regarding societal values. Within these realities a problematic realm is constructed in which specific social groups are excluded or classified as *others.* When this is perceived through the perspective of CDS, it must be argued that these data assemblages not only instigate to represent and construct artificial social realities, but that they are simultaneously being performed. The conception that these realities are constructed with neutral and objective means of quantification and graphic designs must be contested, and the multi-layered entanglement of the data assemblages unpacked. This research offers considerations for the use and representation of open data on CBS platforms as the ODS and its web applications, while simultaneously functions as a starting point to develop sound data practices. We stand at the genesis of the field of CDS as the need for data literacy is ever more eminent. Drawing upon this, the methodological unpacking of the multi-layered data assemblages is especially relevant within the datafication of society.

Future research must entail a more quantitative examination of the extent of this issue, for example by identifying how different types of open data are collected, sorted, classified, categorized

and transformed into graphical form by i.e. the CBS or similar census bureau's. Research examining other similar uses, representations and appropriations of open data by data-driven applications would be interesting to provide insights into the government bodies that publish open data within interactive web applications. Additionally, a more ethnographic user-centred approach, in which the users regarding the interpretations, uses and experimentations with the data, should be examined. Finally, a research on the ODS platform of the CBS through the perspective of CDS could provide interesting findings on how the platform presents data, and which possibilities there are for the user to experiment with the provided data, and the appropriation by third-party applications which make use of the provided open data.

References

Aiello, Giorgia. 2007. 'The appearance of diversity: Visual design and the public communication of EU identity'. In Bain J. and Holland M. (Eds.), European Union identity: Perceptions from Asia and Europe, pp. 147-181

Allen, Rob, and Thijs van den Berg. 2014. 'Serialization in Popular Culture'. Routledge.

Bowker, Geoffrey C., and Susan L. Star. 1999. 'Sorting things out: Classification and its consequences.' Cambridge, MA: MIT press.

Bredekamp, Horst, and Vera Dünkel and Birgit Schneider (eds). 2015. 'The Technical Image: A History of Styles in Scientific Imagery'. University of Chicago Press, 208.

Centraal Bureau voor de Statistiek. 2017. 'Statline Databank as Open Data'. https://www.cbs.nl/en-gb/our-services/open-data/cbs-statline-databank-as-open-data , accessed May 21, 2018.

Centraal Bureau voor de Statistiek. 2017a. 'Organisation'. https://www.cbs.nl/en-gb/about-us/organisation , accessed May 21, 2018.

Centraal Bureau voor de Statistiek. 2017b. 'Wat verstaat het CBS onder een allochtoon?'. https://www.cbs.nl/nl-nl/faq/specifiek/wat-verstaat-het-cbs-onder-een-allochtoon- , accessed May 21, 2018.

Centraal Bureau voor de Statistiek. 2016. 'Welzijn in Nederland'. https://www.cbs.nl/nl-nl/publicatie/2015/45/welzijn-in-nederland, Accessed May 04, 2018.

Centraal Bureau voor de Statistiek. 2016a. 'De Arbeidsmarkt in Cijfers 2016'. https://www.cbs.nl/-/media/_pdf/2017/19/de-arbeidsmarkt-in-cijfers-2016.pdf, Accessed May 06, 2018.

Dalton, Craig, and Jim Thatcher. 2014. 'What Does a Critical Data Studies Look like, and Why Do We Care? Seven Points for a Critical Approach to "Big Data"'. 19 May 2014. https://www.countercartographies.org/critical-data-studies-look-like/.

Dalton, Craig, Linnet Taylor, and Jim Thatcher (alphabetical). 2016. 'Critical Data Studies: A Dialog on Data and Space'. Big Data & Society 3 (1): 2053951716648346. https://doi.org/10.1177/2053951716648346.

Daston, Lorraine, and Peter Galison. 2008. 'Response: Objectivity and Its Critics'. Victorian Studies 50 (4), pp. 666–677. https://doi.org/10.2979/VIC.2008.50.4.666.

Daston, Lorraine, and Peter Galison. 1992. 'The Image of Objectivity'. Representations 40 (October), pp. 81–128. https://doi.org/10.2307/2928741.

Desrosières, Alain. 2013. 'Retroaction: How Indicators Feed Back onto Quantified Actors'. In The World of Indicators, edited by Richard Rottenburg, Sally Engle Merry, Sung-Joon Park, and Johanna Mugler, 132–153. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781316091265.013.

Draude, Claud, Goda Klumbyte, and Pat Treusch. 2017. 'Re-Considering Bias: What Could Bringing Gender Studies and Computing Together Teach Us About Bias in Information Systems'.

Drucker, Johanna. 2011. 'Humanities Approaches to Graphical Display'. Digital Humanities Quarterly 5 (1).

Edwards, Paul, and Christine Borgman and Steven Jackson. 2013. 'Knowledge Infrastructures: Intellectual Frameworks and Research Challenges'. UCLA Publications, University of California.

Foucault, Michel. 1977. 'Discipline and Punish'. Allen Lane, London.

Foucault, Michel. 1980. 'Power/knowledge: Selected interviews and other writings by Michel Foucault 1972–1977'. (C. Gordon, L. Marshall, J. Mepham & K. Soper Trans.). New York: Pantheon Books.

Foucault, Michel. 1991. 'Governmentality', in Burchell, G., Gordon, C. and Miller, P. (eds) The Foucault Effect: Studies in Governmentality. Chicago, IL: University of Chicago Press, pp. 87–104.

Galloway, Alexander. 2011. 'Are Some Things Unrepresentable?' Theory, Culture & Society 28 (7–8): pp. 85–102. https://doi.org/10.1177/0263276411423038.

Geenstijl.nl. 2009. ' Allochtoon-o-meter'. http://www.geenstijl.nl/paginas/mijnburen/?postcode=1064, accessed June 7, 2018.

Gillespie, Tarleton. 2014. 'Algorithmic Culture and Machine Learning: What is an algorithm'.

Gitelman, Lisa. 2013. 'Raw Data Is an Oxymoron'. Cambridge: MIT Press.

Hacking, Ian. 2007. 'Kinds of People: Moving Targets'. In Proceedings of the British Academy, Volume 151, 2006 Lectures, edited by P. J. Marshall, British Academy. https://doi.org/10.5871/bacad/9780197264249.003.0010.

Haraway, Donna. 1988. 'Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective'. Feminist Studies 14 (3): 575.

Harding, Sandra. 2015. 'Objectivity and Diversity: Another Logic of Scientific Research'. University of Chicago Press.

Hassler-Forest, Dan. 2014. 'Game of Thrones: Quality Television and the Cultural Logic of Gentrification'. TV/Series, no. 6 (December). https://doi.org/10.4000/tvseries.323.

Hetherington, Kregg. 2011. 'Guerilla Auditors: the Politics of Transparency in Neoliberal Paraguay'. Durham and London: Duke University Press.

Iliadis, Andrew, and Federica Russo. 2016. 'Critical Data Studies: An Introduction'. Big Data & Society 3 (2): 205395171667423. https://doi.org/10.1177/2053951716674238.

Kennedy, Helen, and 2016. 'The work that visualization conventions do'. Information, Communication & Society 19 (6), pp. 715-735.

Kitchin, Rob, and Claudio Coletta, Gavin McArdle. 2017. 'Urban informatics, governmentality and the logics of urban control'. Luque-Ayala, A. and Marvin, S. (eds) Control Room: Nodes in the Networked City. Routledge, London.

Kitchin, Rob. 2017. 'Introducing the ERC-Funded Programmable City Project | The Programmable City'. Accessed 28 May 2018: http://progcity.maynoothuniversity.ie/about/.

Kitchin, Rob, Tracey P. Lauriault, and Gavin McArdle. 2015. 'Knowing and Governing Cities through Urban Indicators, City Benchmarking and Real-Time Dashboards'. Regional Studies, Regional Science 2 (1): 6–28. https://doi.org/10.1080/21681376.2014.983149.

Kitchin, Rob, and Tracey P. Lauriault. 2014. 'Towards Critical Data Studies: Charting and Unpacking Data Assemblages and Their Work', pp. 3-18.

Kitchin, Rob. 2014. 'The Data Revolution: Big data, open data, data infrastructures & their consequences'. Sage Publications, pp. 32-148.

Lascoumes, Pierre, and Patrick le Gales. 2007. 'Introduction: Understanding Public Policy through Its Instruments—From the Nature of Instruments to the Sociology of Public Policy Instrumentation'. Governance, 20: 1-21. doi:10.1111/j.1468-0491.2007.00342.x

Latour, Bruno. 1991. 'Technology is society made durable'. In J. Law (Ed.), A sociology of monsters: Essays on power, technology and domination, 103-132. London: Routledge.

Latour, Bruno. 2005. 'Reassembling the Social: An Introduction to Actor-Network-Theory'. Oxford: Oxford University Press.

Law, John, and John Urry. 2004. ''Enacting the Social'', Economy and Society, 33: (3), pp. 390-410.

Leidelmeijer, Kees, Gerard Marlet, Roderik Ponds, and René Schulenberg. 2017. 'LEEFBAAROMETER 2.0: INSTRUMENTONTWIKKELING', pp. 35-151.

Leidelmeijer, Kees. 2011. 'Leefbaarheid in balans', pp. 47-113.

Leidelmeijer, Kees. 2008. 'RIVM Rapport 630950002 Kwaliteit van de Leefomgeving En Leefbaarheid', pp. 25-94.

Merry, Sally Engle. 2011. 'Measuring the World: Indicators, Human Rights, and Global Governance: With CA Comment by John M. Conley'. Current Anthropology 52 (S3): 83–95. https://doi.org/10.1086/657241.

Neff, Gina, Anissa Tanweer, Brittany Fiore-Gartland, and Laura Osburn. 2017. 'Critique and Contribute: A Practice-Based Framework for Improving Critical Data Studies and Data Science'. Big Data 5 (2): 85–97. https://doi.org/10.1089/big.2016.0050.

NRC. 2009, December 19th. 'De stad leeft op'. Nrc.nl. Retrieved from https://www.nrc.nl/nieuws/2009/12/19/de-stad-leeft-op-11827412-a161378

Porter, Theodore M. 1995. 'Trust in Numbers: The Pursuit of Objectivity in Science and Public Life'. Princeton, N.J: Princeton University Press.

Porter, Theodore M. 2015. 'The Flight of the Indicator'. In The World of Indicators, edited by Richard Rottenburg, Sally Engle Merry, Sung-Joon Park, and Johanna Mugler, 34–55. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781316091265.002.

Rottenburg, Richard, and Sally Engle Merry. 2015. 'A World of Indicators: The Making of Governmental Knowledge through Quantification'. In The World of Indicators, edited by Richard Rottenburg, Sally Engle Merry, Sung-Joon Park, and Johanna Mugler,Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781316091265.001.

Ruppert, Evelyn. 2015. 'Doing the Transparent State: Open Government Data as Performance Indicators'. In The World of Indicators, edited by Richard Rottenburg, Sally Engle Merry, Sung-Joon Park, and Johanna Mugler, 127–50. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781316091265.006.

Ruppert, Evelyn, John Law, and Mike Savage. 2013. 'Reassembling Social Science Methods: The Challenge of Digital Devices'. Theory, Culture & Society 30 (4): 22–46. https://doi.org/10.1177/0263276413484941.

Ruppert, Evelyn. 2012. 'Population Objects: Interpassive Subjects'. Sociology 45 (2): 218–33. https://doi.org/10.1177/0038038510394027.

Said, Edward. W. 2014. 'Orientalism'. New York: Random House, Inc.

Smit, Alex and Gerwin van Schie. Forthcoming. 'The Politics of the Dutch Leefbaarometer: A Critical Data Studies Approach'.

Spivak, Gayatri. C. 1988. 'Can the subaltern speak?' In C. Nelson, & L. Grossberg (Eds.), Marxism and the interpretation of culture. Urbana: University of Illinois Press, pp. 271-313.

Stanford History Education Group. 2016. 'Evaluating Information: The Cornerstone of Civic Online Reasoning'. Stanford. https://sheg.stanford.edu/upload/V3LessonPlans/ExecutiveSummary 11.21.16.pdf , accessed June 11, 2018.

Strathern, Marilyn. 2000. 'The Tyranny of Transparency', British Educational Research Journal 26(3): 309–321.

Teo, Thomas. 2010. 'What is Epistemological Violence in the Empirical Social Sciences'. In Social and Personality Psychology Compass 4-5(5): 295-303. https://doi.org/10.1111/j.1751-9004.2012.00265.x

Vertesi, Janet, and Paul Dourish. 2011. 'The Value of Data: Considering the Context of Production in Data Economies'. ACM Press. https://doi.org/10.1145/1958824.1958906.

Wohlers, Tony E., and Lynne L. Bernier. 2015. 'Transformation of Local Government in the Digital Age'. In Setting Sail into the Age of Digital Local Government, 29–36. Public Administration and Information Technology. Springer, Boston, MA. https://doi.org/10.1007/978-1-4899-7665-9_3.

Appendix

1. Transforming data into indicators and graphic design

In order to transform data into usable input for data-driven instruments, a minimum of four steps have to be taken into account (see figure 1). The four steps consist of: 1) create equivalences and/or relations within data; 2) develop (data) categories; 3) code phenomena into (data) categories; 4) construct indicators representing the processed data. Within this example an assumption is made that all data residing in the databanks is cleaned, accessible and ready to use by the various systems. An example of domestic child abuse is used to make the process more clear. As this is a sensitive subject, I am aware of the need to be cautious and use it solely to demonstrate how interpretative and pragmatic the process of transforming open data into indicators is.
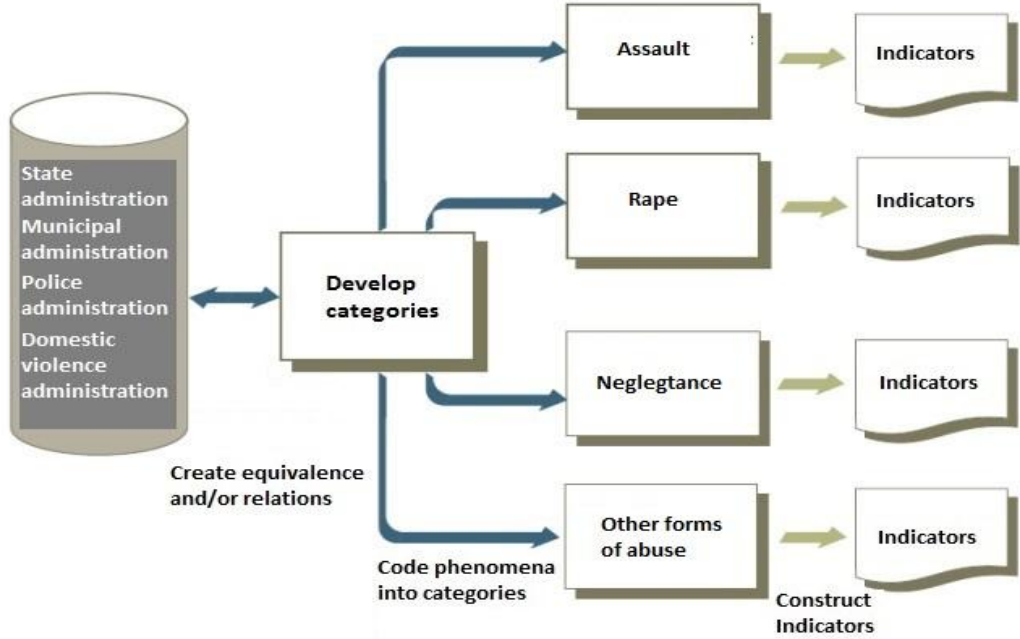
Figure 20: Constructing indicators

The first step within the process is to *create equivalences across different data formats*: data formats have to be homogenized to group and relate them to each other. In order to do so, commonalities and relatable traits have to be established. Differences and heterogeneous traits within the data have to be modified or ignored. This becomes clear with the following example: if a researcher wants to perform a quantitative research on domestic child abuse, he/she needs to achieve interchange-ability within a broad spectrum of forms on domestic violence. These forms could include neglecting, intimidating, injuring, attacking, stealing, humiliating, confining food and financial support. All these

could be attributed to physical, emotional and mental forms of domestic child abuse, while they all are various actions and embed different meanings, experiences and effects within a domestic sphere. It is essential for the researcher to perceive these actions as manifestations of the same, so that it could be quantified and calculated. A manner has to be constructed in which the distinct actions are made equivalent in relation to one topic, in this case domestic child abuse.

The second step is to *develop (data) categories and classification schemes*: the data categories have to be organized with a classification scheme, in order to divide all data in certain classes which are mutually exclusive (i.e. a taxonomy) (Ruppert 2015). Constructing such categories implies the creation of a spectrum in which a progression is made from the individual act to the general, with the help of consistent equivalent classifications. Returning to the former example, domestic child abuse could be categorized using official legal categorizations as threat, assault, rape and others, or it could be categorized on a continuum going from less to more severe abuse. Both these approaches enable to divide an action into a category, and they both have certain values in examining various dimensions of the same manifestation. Classifications often reflect the values, norms, and concerns within societies, while simultaneously affecting and shaping these subjects (Rottenburg and Merry 2015).

The third step is to *code phenomena into (data) categories and classifications*: individual, random and various phenomena's are coded into categories constructed in the former step. This process of encoding relates to the attribution of data to a specified class. Again, this is a highly interpretative process which ignores or consciously neglects certain details of phenomena, while foregrounding others. The appearance that 'factual' 'given' statistical data are merely quantified and categorized, construct a simplified conception of complex phenomena, as it becomes 'objectively' real after being encoded. The categories become naturalized over time, as they are used more and more by various actors (Porter 1995).

The fourth step *uses categories to construct indicators*: as the categories are constructed with according taxonomies, these can be attributed to specific indicators. Indicators could be defined as measurement units or as a mathematical aggregation of multiple measurements. I.e. obesity cannot be discovered by only measuring weight. It has to be related to data on age, height and sex. Another example is the broadly known indicator of wealth in a country: the Gross Domestic Product (GDP). This indicator uses various open government data to determine the wealth of a country within a specified year, but is questionable and often criticized by being a weak measure of wealth as it strongly focuses on very specific socio-economic aspects in society while neglecting others (Desrosiéres 1998; Porter 1995; Speich 2013).

These steps are all highly interpretative processes, while data categories, taxonomies, indicators and data visualizations are all commonly regarded as 'objective' quantification instruments

of societal phenomena. These processes are not merely driven by the will to simply quantify, classify and visualize societal aspects, but foremost what certain actors want to know, and especially by practical concerns relating to finance, technology, time and other resources, and by political, cultural, social, economic and historical reference frames which affect and shape these epistemic practices (Ruppert 2012).

After the construction of indicators, designers of data-driven instruments generally divide the indicators into dimensions generalizing those indicators within a broader frame of socio-cultural contexts. Returning to the example of domestic child abuse, you could think of dimensions stating physical abuse or mental abuse. These dimensions are often (implicitly) embedded within the data visualizations. It does depend on the context of the instrument how explicitly these dimensions are represented and visualized.
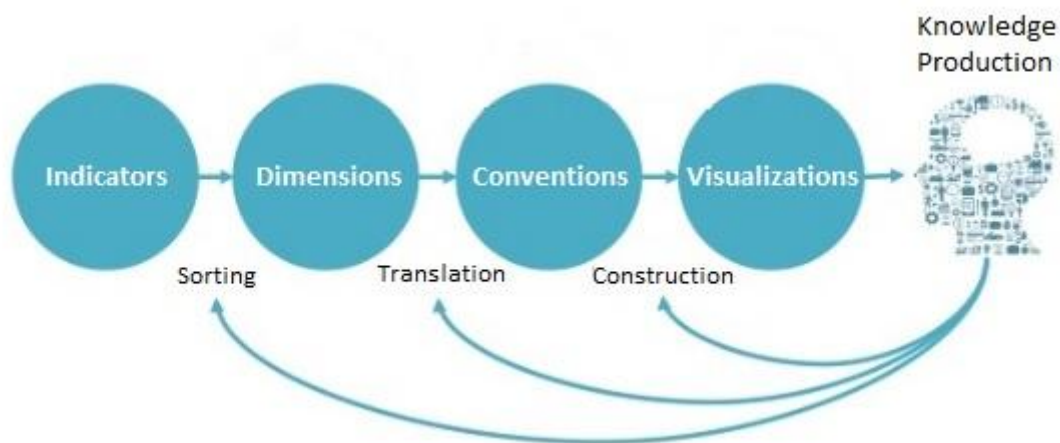
Figure 21: Knowledge production process

Each element affects the epistemic process in a different manner. When indicators are sorted and grouped into several dimensions, the context of the initial data disappears and various indicators are grouped together while the explicit relational and causal factors are missing. Next, the dimensions are translated into conventions, they get objectified and embedded within the visualization practice itself. Now the visualizations are constructed upon the former iterations and the initial data with according indicators are (almost) invisible in visualized format. Finally, the public looks at these visualizations and interprets them within a personal frame which is highly affected by the multi-layered elements in the epistemic process (Bredekamp et al. 2015, p. 2).

2. Discursive Interface Analysis: The Walkthrough Method

The first part of 'The Walkthrough Method' focuses on examining the vision of an application. Within this vision the purpose, target user base, and scenarios of use are points of inquiry. These elements

help establishing an app/tool's environment of expected use. These are mostly communicated through technology's organisational materials and its interface (Light et al. 2016). The second part of 'The Walkthrough Method' is called the technical walkthrough and is the method's central data-gathering procedure. It involves the researcher engaging with the applications interface, working through screens, tapping buttons, and exploring menus. Walking through the application applying an analytical eye to the process consisting of three phases being that of acquiring, registering and entering the application, use of the features, and discontinuing use of the application. The registry and entry phases of the expected use and the governance are generally strongly communicated. Whereas in the analysis of the use, the method focuses on the functionality, options, and affordances, but also the activity of the user. The final phase involves attention to the applications materiality and the physical interactions encouraged, it involves the analysis of the actions it requires a user to undertake, and other affordances like buttons, icons, and so forth, and finally it requires imaging how users would perceive these as affordances or constraints. The difference from this method to it instructional predecessors, is its use. It is not to instruct or analyze a user, but to analyze the application itself, and to illuminate the material traces of designer intentions. Thereby to critically examine the workings of an application as a socio-technical artefact, or within the perspective of this research, as a *data-assemblage* (Kitchin and Lauriault 2014)*.

2.1 The Geluksmeter

2.1.1 Scenario of use

When visiting the homepage of the Geluksmeter, you see a map of the Netherlands with the circular visualization of the Geluksmeter, and a button with the question "hoe gelukkig is Nederland? Klik hier!" (see fig. 21). The circular bar chart is accompanied with a cloud with rain and thunder, a cloud with rain, a cloud, a cloud with sun and a sun. This symbolically shows the value of the bar chart ranging from negative towards positive.

Figure 21: Geluksmeter Homepage

The only interactive option for the user is to click on the button "hoe gelukkig is Nederland? Klik hier!". When clicking on this button the average 'geluksscore' of the Netherlands is depicted. This is a 7,1. Additionally, the question "hoe zit dat in jouw provincie?" is asked, which means that the user can see what the average happiness score is for his/her province. Clicking on this question is the only option the user gets to move forward within the interface. Next, a map of the Netherlands is visualized, on which the user can click to select a province to see its average 'geluksscore' (see fig. 22).

Figure 22: Geluksmeter Location Selection

The province is built upon numerous small black dots, which represent people within that area. When a province is selected, the average score for that area is visualized. Next, you have to choose your gender to progress (see figure 23). A binary is established whereas the user can choose between male or female.

Figure 23: Geluksmeter Gender Selection

When a gender is chosen, you have to further specify your personalized happiness score by selecting an age group in which you are situated. The age groups consist of: 18 – 24, 25 – 44, 45 – 64, and 65+ (see fig. 24). It is not stated why an age group below 18 years of age is not incorporated within the application. The only interactive function is the "I" button on the top right corner of the screen.

Clicking on this button shows a text which elaborates what the user has to do to continue in the system.
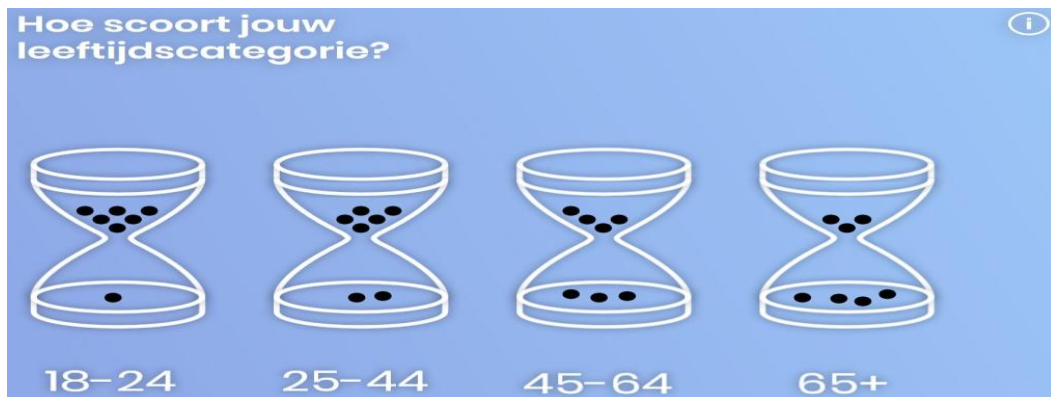
Figure 24: Geluksmeter Age Selection

When selecting a specific age group, you have to further specify your personal characteristics by selecting your level of education. You can choose from: high school, VMBO, HAVO, VWO, MBO, HBO, and WO. After this selection an average is given based on your personalization variables, and the eight dimensions are visualized within a circular graph. If you use your mouse to hover above the visualization, different scores are presented based on the eight dimensions. Finally, you can go on to the questionnaire where you have to answer eight questions based on the eight dimensions on a scale from one to ten. After answering these questions your personal score is calculated and you can compare these with the average scores within the eight dimension (see fig. 25).
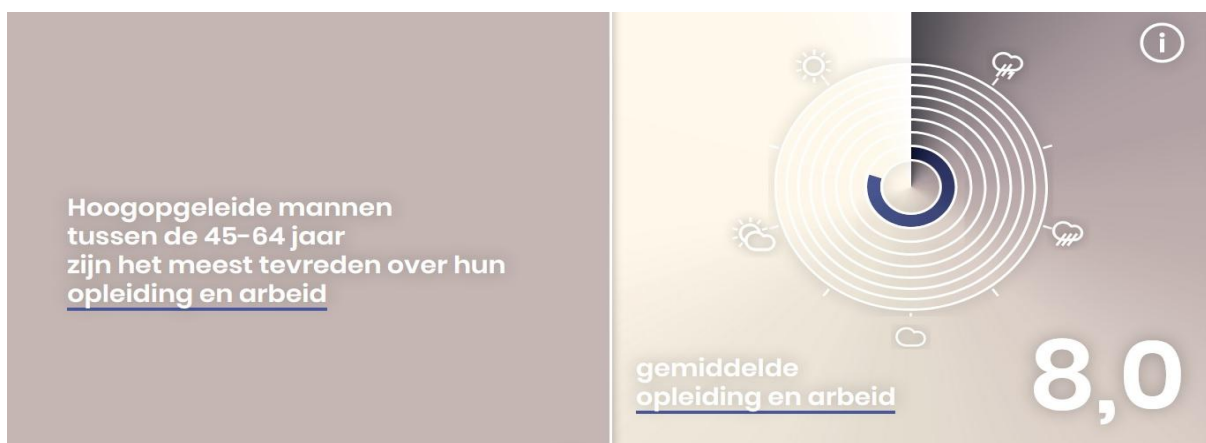
Figure 25: Geluksmeter Circular Bar Chart Representation

Here the dimension education and work has been selected. You can see that the average number for this dimension consists of a 8.0. This is the average for high educated men between the age group of 45-64. Men within this age group are most satisfied about this dimension. If you select female as a

gender, and the same age group and education level, you can notice that woman are most satisfied within the liveable surroundings dimension (see fig. 26).

Figure 26: Geluksmeter Circular Bar Chart Representation with Dimensions

No further information is given how these dimensions are calculated or visualized. The user is not given an option to gain deeper insights into the calculation or representation of the used data. The interface is thus verily simplistic in its functions, as the user only has a few interactive options in each screen which it can alter and use to navigate through the system.

2.1.3 Governance

The governance consists of how the Geluksmeter seeks to manage and regulate user activity in order to sustain its operating model and fulfil its vision (Light et al. 2016). It can expand from managing user activity to enforcing certain values and norms. It is often enacted through the Terms of Service (ToS), while the Geluksmeter is lacking such ToS. This means it is hardcoded into the features of the interface. This can be seen by focusing on the strong regime and limited possibilities the user has to manoeuvre through the interface. The interface forces you to choose if you are a male or female, in which age group you are, what education level you have, and where you live. You do not have the freedom to provide any other answers besides those which are provided by the interface. This could be perceived as a way to enforce certain norms and values, which are present within the report 'Welzijn in Nederland' and the Geluksmeter, on the user. If you for instance are a transgender and are not situated within the represented variables, you do not have additional options to personalize the Geluksmeter and thus cannot calculate a personalized 'geluksscore'.

2.1.4 Operational Model

The operational model examines the business strategy and revenue sources of the Geluksmeter. The Geluksmeter is a state funded application, constructed by the CBS (CBS 2016). It has no commercial interests, what fits within the purpose to inform the public on statistical data from the CBS. Furthermore, when visiting the web application no cookies or trackers are installed on the user's computer, which tells that the business strategy and revenue source is not aimed at collecting user data and/or data for commercial purposes. This fits in line with the purpose of the application and motivation according to the official documentation.

2.1.5 Technical features:

Software: Adobe Flash Player

Version: 30.0.0.113 (64 bit)

Supported devices: portables (laptop, macbook, ipad, kindle, mobile phones); pc, imacs

Updated: last update in 11-2016

2.2 The Leefbaarometer

2.2.1 Scenario of use

When visiting the homepage of the Leefbaarometer, you see a small geographical map of a municipal area within the Netherlands, with nine colours (ranging from dark red to dark green) which score an area based on the liveability data (see figure. 27).

Figure 27: Leefbaarometer Homepage

A welcoming text is accompanying the map concerning a small introduction on the workings of the Leefbaarometer and its features. The user can click on the buttons "home", "kaart", "tabellen", "nieuws", "open data", "publicaties", and "help". Furthermore the user can adjust certain parameters and variables concerning the measurement and visualization of the liveability score within an area. The user has the choise between the indicators of the present liveability situation, or the liveability development within a specified area. Next, a scale can be chosen to be represented on the geographical map. The user can choose between a map on a municipal scale, zip code scale, neighbourhood scale, and grid-map scale with grids consisting of 100 x 100 square meters. Next, a

time period can be selected ranging from 2014 to 2016. Furthermore a reference frame with certain boundaries can be chosen ranging from municipal boundaries, zip code boundaries, neighbourhood boundaries, grid-map boundaries and boundaries which were established in the first version of the LM. Finally the user can search for a specified location within the Netherlands, or use a GPS tracker to detect the current location of the user and visualize it on the geographical map. Besides these features, the user can select news articles and publications relating to the Leefbaarometer. When the user clicks on the button "kaart" which is translated as "map", the geographical map of the Netherlands is visualized, as shown in fig. 28.
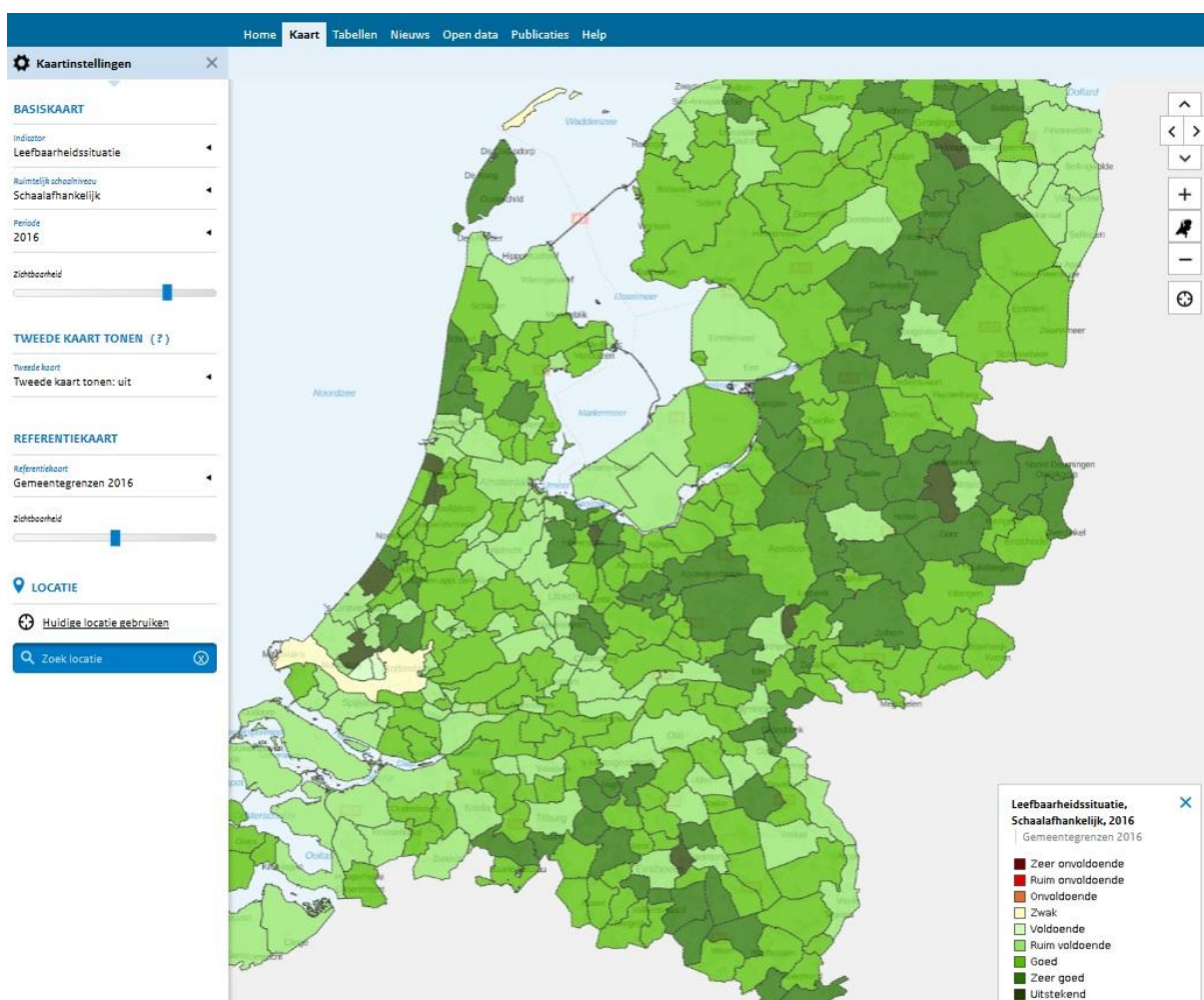
Figure 28: Leefbaarometer Geographical Map

The user has the option to navigate through the map using cursors on the top right side of the screen. The user can zoom in, and out following the specified scales, and can read the various colours with attributed scores in the lower right corner of the screen. The colours vary from dark red (very insufficient score) to dark green (outstanding score). On the left of the screen the user can select various options, parameters and variables which can alter the visualizations on the map. Similar parameters can be selected as in the previous screen, from the liveability situation and development

to the visualized scale and time frame. An additional slide-bar is shown whereas the user can shift the transparency of selected criteria. The whole preferences window can be shut so the user has a larger visualization window on which the geographical map is depicted. When a user zooms in on the map (in this example I have zoomed in on the municipality of Utrecht), the liveability scores are given on a zip code scale with official boundaries as documented in the GBA databank (see fig. 29).
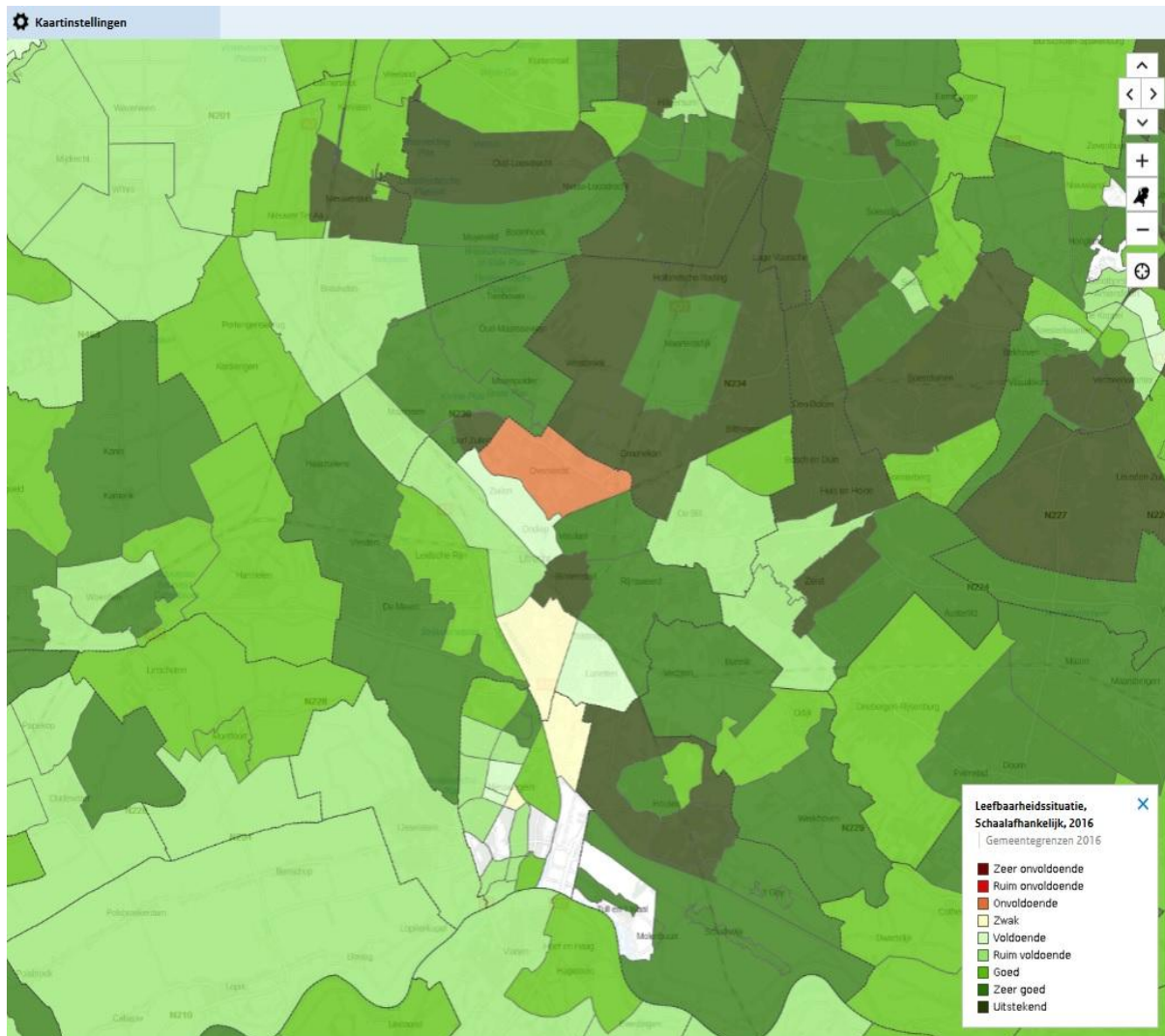
Figure 29: Leefbaarometer Geographical Map Municipal Scale

Within Utrecht most of the areas are scored in a positive manner, only 'Overvecht' and 'Kanaleneiland' are scored negative. The blank areas are not scores because less than the minimum amount of forty people reside within the specified area. Due to privacy regulations this minimum has been established so that the given scores cannot be personalized or attributed to specific inhabitants of a scored area. When the user zooms in even further a grid-map is shown where a more precise picture is given which areas are scored in what manner (see fig. 30).
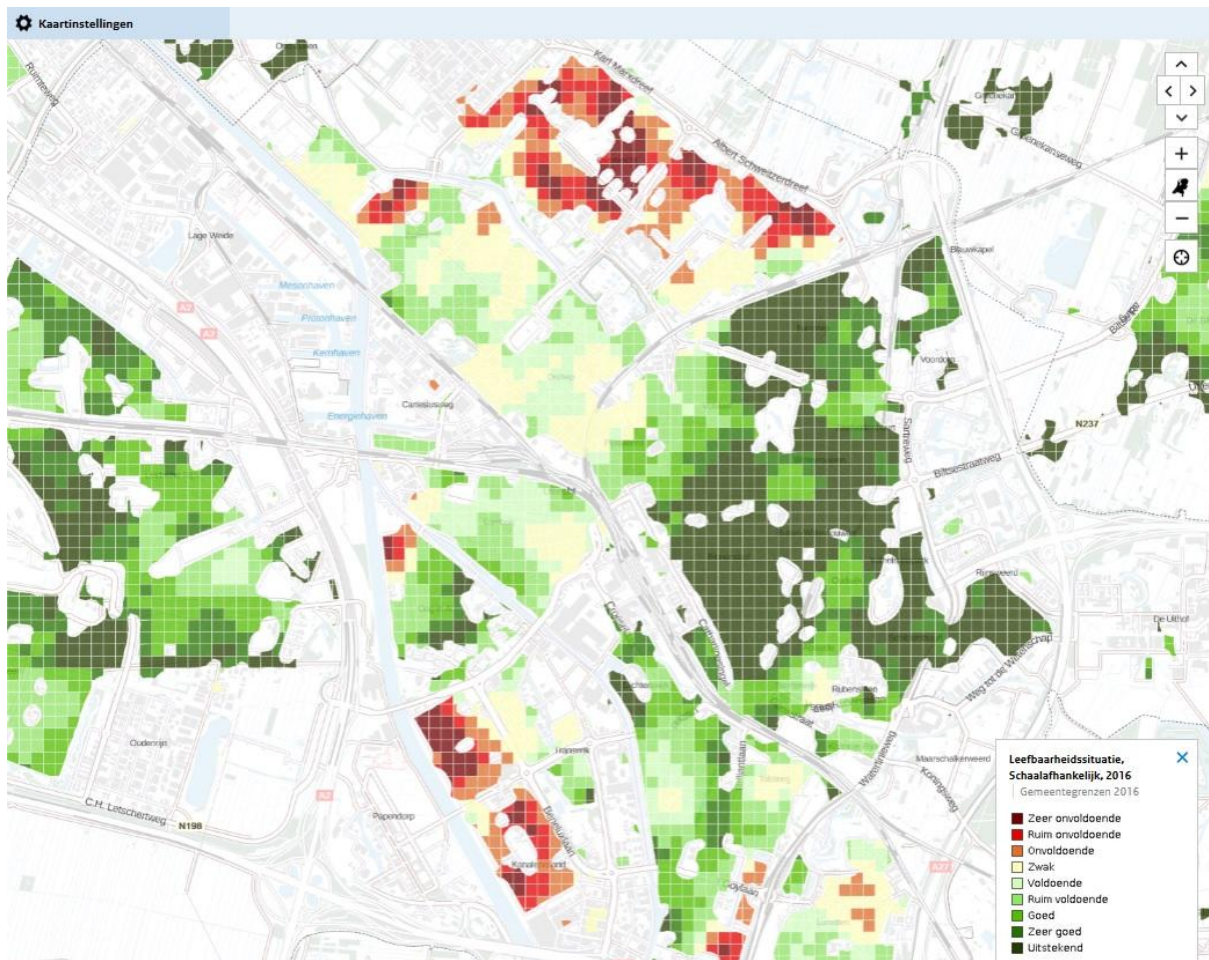
Figure 30: Leefbaarometer Geographical Map Grid-Map

On this image the map is scaled on a grid-map level where the given score can vary according to each grid of 100x100 square meters. A big difference can be noticed as far less areas are scored than in the previous image on a higher scale, due to the minimum inhabitants limit. Furthermore some grids located in 'Overvecht' and 'Kanaleneiland' are scored even more negative (dark red) than on a higher scale. The inner city of Utrecht is scored very positive in comparison with 'Overvecht' and 'Kanaleneiland'. Zooming in even further does not change the scores, it only visualizes the names of streets on the map (see fig. 31).
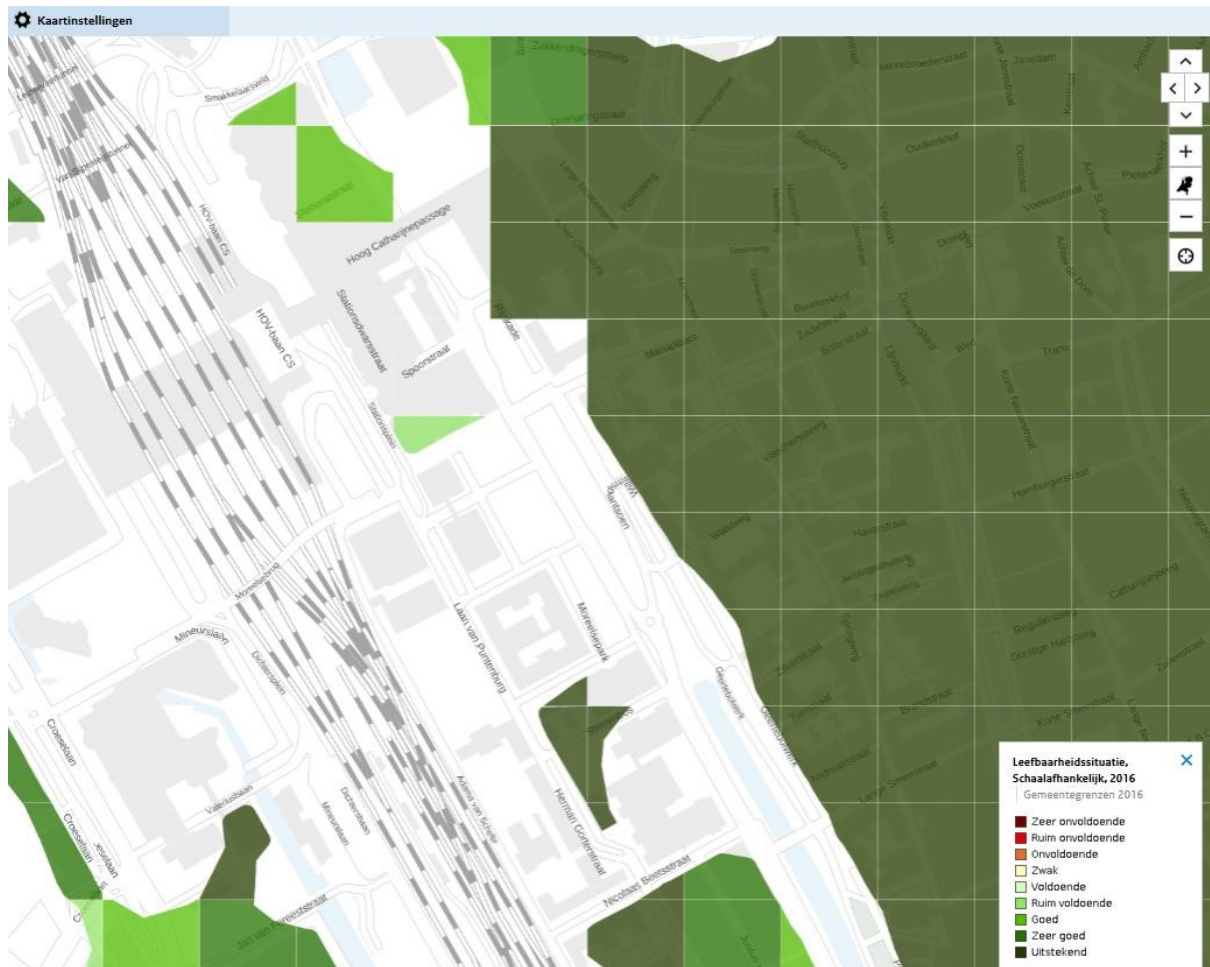
Figure 31: Leefbaarometer Geographical Map Grid-Map Zoomed In

The grids are scored the same, with only additional information concerning the street names and geographical points as canals, buildings, roads, tracks and various features become better visible. The user cannot zoom in further, so the only interactive option left is clicking on one of the grids. When the user clicks on one of the grids an information window is shown with a identification number of the selected area, in which municipality it is located, the selected time frame, and the given score by the Leefbaarometer (see fig. 32). Furthermore a bar chart is visualized with all of the five dimensions (buildings, population, facilities, safety and physical surroundings). Furthermore a total score of deviation is visualized as the top bar chart. The bar chart horizontal scale ranges from -0.3 to 0.3, which depicts a measure in percentages of -30% to 30% deviation from the average or set norm.
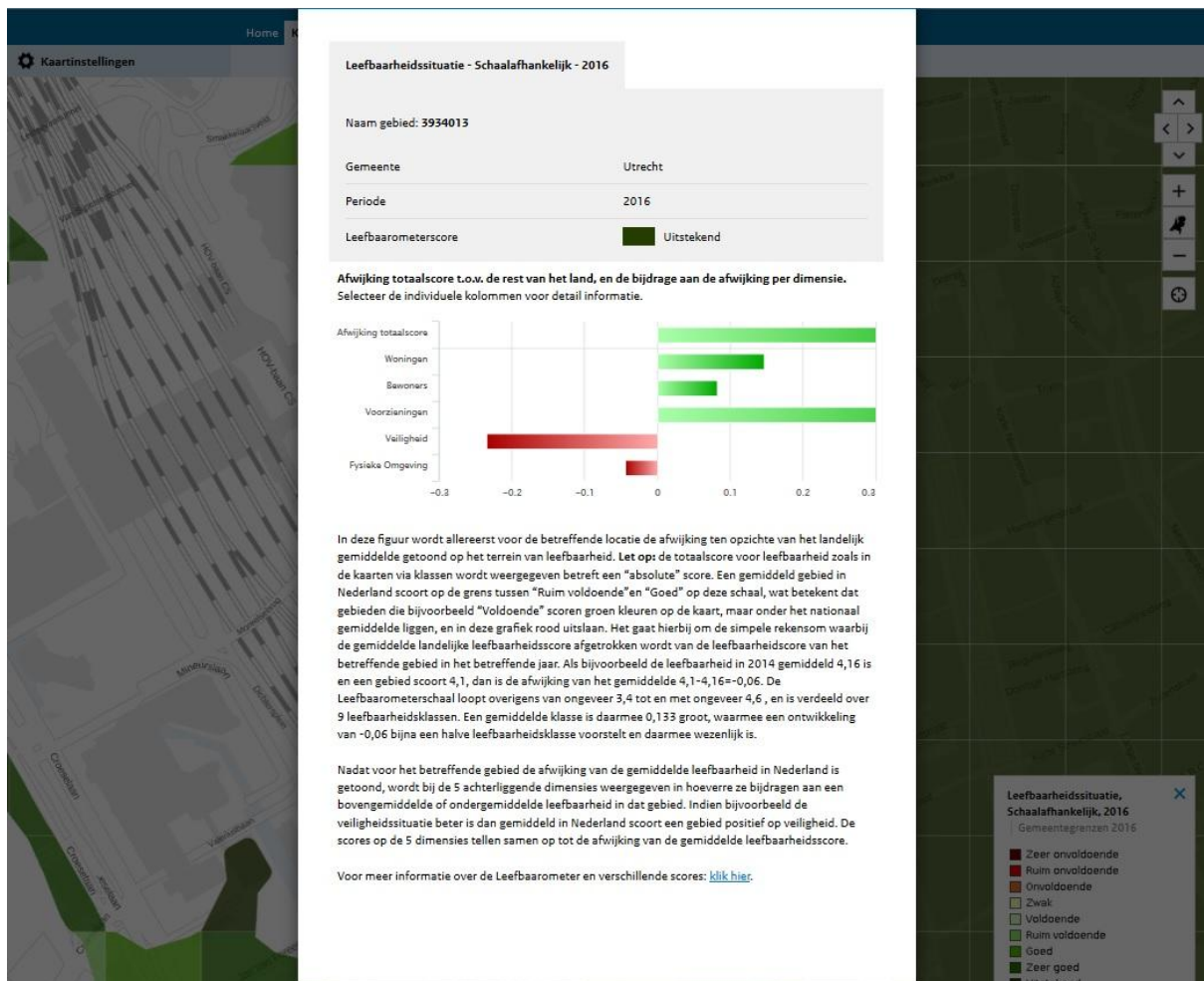
Figure 32: Leefbaarometer Geographical Map Additional Information

Beneath the bar charts additional information accompanying the measurement and calculation of the liveability score is described. The information concerns how the liveability measure is calculated according to the five dimensions, and how it relates to an average number. An elaboration is made on what the different scoring measures mean, and which colours are attributed. It describes that the scoring scale ranges from 3,4 to 4,6 and is divided into 9 liveability classes, in which an average class has a value of 1,33. In this measure a development of 0,06 means relates to a half of a liveability class. After a deviation from the average in that area is established, each of the five dimensions show its deviation from the average numbers. If the user wants additional information on this calculation and given scores the user is forwarded towards the help page (see fig. 33).

Figure 33: Leefbaarometer Help Page

On this help page the user can look up information regarding subjects as what the Leefbaarometer is, why it has been developed, what liveability means and how it is defined, how the liveability score gets calculated and measured, what dimensions and indicators are at the base of the liveability calculation, what the various represented scales are, what the differences between the first and second versions of the Leefbaarometer are, and which publications concerning the Leefbaarometer and its liveability measure have been publicized. If the user wants any additional information the user is forwarded towards official documentation of the CBS. More information can be found if the user clicks on the "tabellen" button in the top of the screen (see fig. 34).

Figure 34: Leefbaarometer Tables Page

The user can make a selection based on the liveability situation or development of municipalities within the time frame of 2002 up till 2016. A table is shown where per year a score of liveability within that municipality is given. This ranges from "ruim onvoldoende" (very insufficient) to "zeer goed" (very good). The user cannot click on the table to request additional information or see how the scores are calculated within that specific time frame and municipality. The only interactive feature of the table is the blue with white arrows next to the represented year on the top of the table. Clicking on these arrows filters the table according to worst – best scores, or best – worst

scores. When the user moves towards the next page, the button "nieuws" is clicked. Here one article is shown originating from 11-12-2017 relating to liveability within the Netherlands (see fig. 35).

Figure 35: Leefbaarometer News Page

The user can click on the article to see its content, which regards how municipalities can use the Leefbaarometer to proactively notice decline in their liveability level with the use of the Leefbaarometer. A visualization is shown of how the liveability measure and use of the Leefbaarometer can help examine this change in liveability level within a municipality. Clicking the button "open data" guides the user to the next page. Clicking on this button shows the user information on the use of open data by the Leefbaarometer (see fig. 36).

Figure 36: Leefbaarometer Open Data Page

The text elaborates on the used open data and where, and how it can be accessed and viewed. It is noted that a "Creative Commons" license is attributed to the datasets, which means that the data can be used by various actors for all means. The user can download the scores and shapefiles relating to the used open datasets. Additionally, the user can use the "Web Map Service" or the "Web Feature Service" to use the visualizations on the geographical map within its own software, website or system and appropriate it accordingly. If the user would like to download the used open data as an Excel file, the user has to click on the provided link under "Download scores". The user then gets forwarded towards the Open Data Platform of the Dutch government, where the used datasets within the Leefbaarometer can be accessed, viewed and downloaded (see fig. 37).

Figure 37: Leefbaarometer Open Data Page With Datasets

A description is shown of the dataset, the status of the dataset, what the dataset contains, how it is licensed, who can access it, the ownership of the data, and various tags relating to the dataset. The dataset can be downloaded in XLS format, and information is shown when the dataset has been checked for the last time (it shows that this was done for the last time at 25-06-2018). Within the dataset menu six buttons are shown, consisting of "datasets", "organisaties", "groepen", "dataset aanmelden", "toelichting data" and "statistieken". Within the "datasets" window the user can download various XLS files consisting of datasets ranging from 2002 up till 2016 regarding the liveability measure of municipalities within the Netherlands. The user can furthermore click on the button "overige", "Documentatie" and "Metadata" which mean "others", "documentation" and "metadata" (translated by author). When downloading a dataset three different XLS files are downloaded within a .ZIP packaged file. The three files are: "file 1 dimensiescore_gemeente (ontwikkeling)", "file 2 dimensiescore_gemeente (stand)", and "file 3 score_gemeente". Within the Excel files rows with values attributed to the five dimensions can be found, divided into the present state (file 2), development (file 1) and national averages (file 3). No metadata is accompanied how these values are calculated or attributed to the specific dimensions. To look into the metadata the

user has to click on the "metadata" button within the open data platform. The metadata is then shown within a table consisting of five sections: "A: organisatie", "B: verplichte velden", "C: vindbaarheid", "D: kenmerken data" and "E: aanwijzingen voor hergebruik" (see fig. 38). These are translated (by author) as A: organisation, B: constrained fields, C: ability to find, D: characteristics data, and E: directions for re-usability.

Figure 38: Leefbaarometer Open Data Page With Metadata

Within section A, the table shows metadata concerning the ownership of the data, contact information of the owners, the facilitator of the data, contact information of the facilitators, and to which catalogue of the governmental database the dataset belongs. In section B the table shows information regarding the status of the dataset, the license, the language, and the last time the dataset has been configured. In section C information is shown regarding the theme of the dataset.

In section D information is shown on the location which the data is related to, the timeframe, the update cycle, the linked data scores, if the dataset is a "high value" dataset, if the data is "reference data", and if it contains or relates to the "basic register". Section E shows information on the accessibility of the dataset, constrains on the re-usability, if the dataset is closed and the specifics of the data. These last three subjects are left bank. Additionally the user can request specific datasets by clicking on the "dataverzoek doen" button, or give feedback on the datasets by clicking on the "feedback" button.

Going back to the Leefbaarometer interface, the user has one more window which has not been elaborated yet, showing additional information on the application such as the disclaimer, contact details, technical features, and the use of cookies and privacy statement. Here the user can get contextual information regarding the Leefbaarometer application and its creation, motivation and governance. These elements will be described in the following section.

2.2.3 Governance

The governance consists of how the Leefbaarometer seeks to manage and regulate user activity in order to sustain its operating model and fulfil its vision (Light et al. 2016). It can expand from managing user activity to enforcing certain behaviour. It is often enacted through the Terms of Service (ToS) and possibilities the user is given to manoeuvre through the interface, what has been shown in the previous part. The governance is most explicitly described in the disclaimer and ToS accompanying the application, as such these will be examined. Regarding the ToS on privacy and cookies it states that grids on the grid-map visualizations do not specify a score if less than one hundred people live within that area. This is done because of the fact that the given score cannot be specified to a specific person. Furthermore it states that de boundaries as shown within the map are not solid, i.e. two grids next to each other can overlap, but for the ease of the visualization a hard boundary is represented. The grids are diffuse, which means that the user cannot select one grid and say that this area has a specific liveability level while another grid next to the specified area has a whole different liveability level.

Next, the use of cookies and collection of user data is regarded. It is noted that the Leefbaarometer uses three types of tracking cookies to monitor user activity. These cookies consist of: "cookie_pk_id" (tracks if the user is new or did already visit the application before); "cookie_pk_ses" (tracks which pages the user has viewed); and "cookie_pk_ref" (tracks which sites the user visited before it entered the application). It is stated that this information is only used to enhance the Leefbaarometer application and it is not sold or given to other actors. It is made explicit that these cookies do not affect the privacy of the user, and thus are not bound by the Dutch Telecommunications law in which a website has to ask the user if it accepts the placement and use of

cookies and tracking of its internet activity. Additionally, the Leefbaarometer uses web statistics with the data analysis software "Piwik". The platform of the government (Rijksoverheid.nl) analyses the collected data to enrich its applications and open data platform. It is again made explicit that this data is only used for the benefit of the platforms and websites of the government and its bureaus.

In the disclaimer an elaboration is given on why and how areas are not scored due to less than the established minimum inhabitants of the area. Additionally, the purpose and hypothetical use of the Leefbaarometer are described as to inform the public on the liveability of areas within the Netherlands, and the development of liveability within municipalities. It is stated that the Leefbaarometer should be used within municipal policy making to give a sound representation of liveability within specified areas, and notice the development of the liveability. This could be used to create or change policy regarding the change in liveability measure. The Ministry of Internal Affairs uses the Leefbaarometer to check if its policy is working as planned or something has to be changed. This was an assignment by the Dutch congress "Tweede Kamer", that wanted every two years to be informed on the Dutch level of liveability. Additionally it is described that the Leefbaarometer is used as an instrument to provide insights and arguments based on given numbers on liveability regarding the allowances to give concerning the battle against urban problems by municipalities. Next, information is given on publications regarding the Leefbaarometer, and where the user can download these publications. Additionally, a colophon is added which states the parties that helped construct the Leefbaarometer, consisting of: Ministry of Internal Affairs, Rigo Research and Advice Centre, Atlas voor Gemeenten (Utrecht), In Fact (Amsterdam), and OTB (Delft). The design and realisation of the interface and application has been done by 2CoolMonkeys (Utrecht), and Geodan (Amsterdam).

Concluding, an elaboration is made regarding the construction of the Leefbaarometer 1.0, and the changes which occurred in the second version. The first version was constructed with the help of forty-five liveability experts from the scientific field and the governmental field, which discussed and reported on how and why a liveability measure should be calculated and scored in an application like the Leefbaarometer. While the second version of the Leefbaarometer came into being with an analysis of the first version by an independent commission reporting their findings to the Ministry of Internal Affairs, which consisted of the following people:

- Wim van der Zanden, Gemeente Rotterdam
- Thijs van der Steeg, Gemeente Almere
- Marian Huisman, Gemeente Almere
- Janine Meesters, Gemeente 's-Hertogenbosch
- Harry ten Caten, Gemeente Eindhoven / VSO Vereniging voor Statistiek en Onderzoek
- Huub Hanssen, Provincie Groningen
- Jacqueline Gommans, Gemeente Dordrecht

Based on their findings changes within the Leefbaarometer 2.0 where made within the statistical model, indicators, dimensions, values, simulation models, regression analyses and use of data. Now that the governance is established, the analysis move towards the operational model of the Leefbaarometer.

2.2.4 Operational Model

The operational model examines the business strategy and revenue sources of the Leefbaarometer. The Leefbaarometer is a state funded application, constructed by the Ministry of Internal Affairs, CBS, RIGO, Atlas in 2008 (CBS 2016). It has no commercial interests, what fits within the purpose to inform the public on statistical data from the CBS. Furthermore, when visiting the web application three cookies are installed on the user's computer without the purpose to collect user data and sell it or give the data to other actors. The data is solely used to enrich the website, platforms and applications, which tells that the business strategy and revenue source is not aimed at collecting user data and/or data for commercial purposes. This fits in line with the purpose of the application and motivation according to the official documentation.

2.2.5 Technical features:

Software: Adobe Flash Player; Piwik

Version: 30.0.0.113 (64 bit)

Supported devices: portables (laptop, macbook, ipad, kindle, mobile phones); pc, imacs

Updated: last update in 11-01-2018

2.3 The Uitkeringsmonitor

2.3.1 Scenario of use

When entering the application the user gets to see a Venn diagram where two circles are already filled with social welfare types as "bijstand" and "WW", which mean social benefits and unemployment support (translated by author). The third circle is still blank as the user can click on it and select a social welfare type it wishes to incorporate in the visualization (see fig. 39).
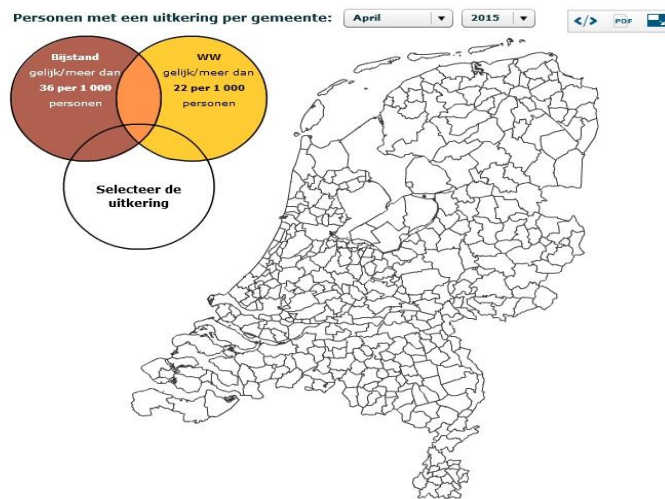
Personen met een uitkering per gemeente

Figure 39: Uitkeringsmonitor Venn Diagram and Geographical Map

When the third circle is selected the geographical map of the Netherlands is filled with the colours attributed to the social welfare classifications within the Venn diagram. The user can select the following kinds of social welfare/benefits within a Venn diagram to be visualized on the map: law for the disabled (AO), law on disability allowance (WAO), law on work and income for the capability to work (WIA), law on work and aid to work for disabled young adults (Wajong), law on unemployment (WW), and the law on general welfare (Bijstand) (translated by author). A colour can then be assigned to each category of social welfare/benefits recipients whereas it is visualized with a comparing measure of social welfare/benefits recipients of one per thousand inhabitants within a specified area (see fig. 40).
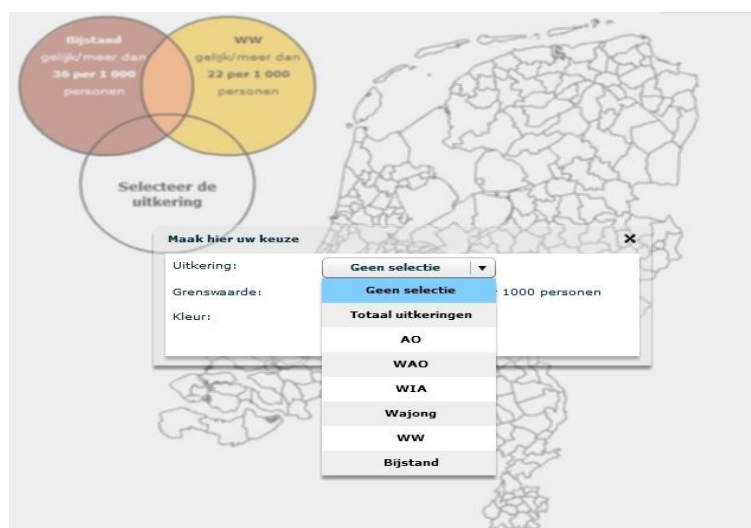

Figure 40: Uitkeringsmonitor Venn Diagram and Welfare Selection

The following map shows the incorporation of the selected welfare types in the geographical map of the Netherlands. The Venn diagram makes the colours mix and appear as different multiple colours on the map. The user has to notice carefully which mixed colours are attributed to which mixture of social welfare/benefits recipients.
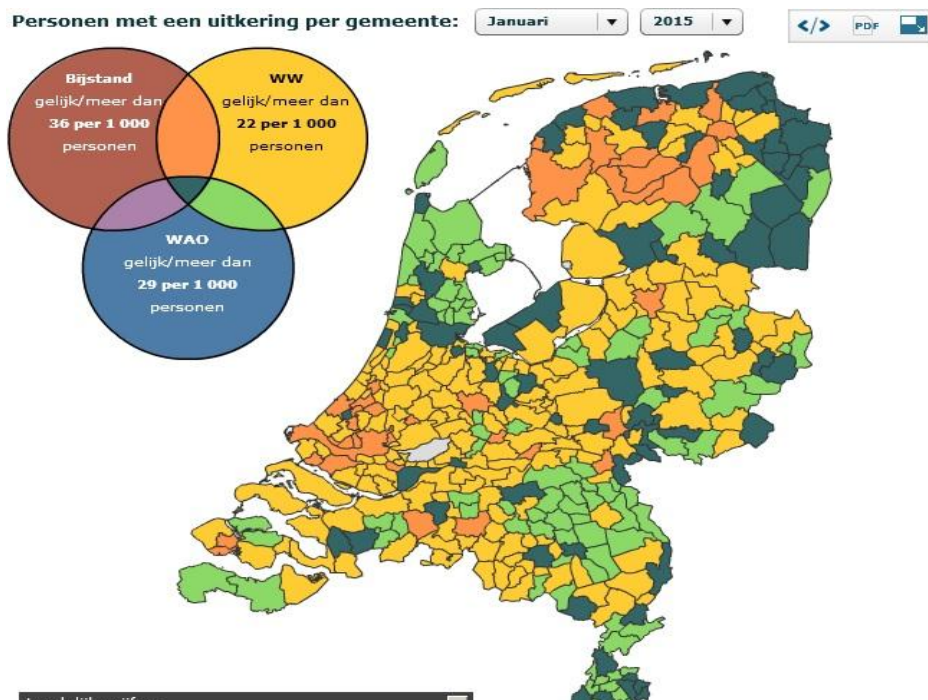

Figure 41: Uitkeringsmonitor Venn Diagram and Welfare Selected Map

In this example the user selected dark red for the general welfare recipients, yellow for the unemployment welfare recipients and blue for the disability allowance recipients. The Venn diagram mixes these colours to represent a mixture of these recipients on the map. Above the map the user can select a time frame of months and years, ranging from 2010 to 2015. Futhermore the user can make the application go to full-screen mode, download the visualization of the welfare types on the map in PDF format, or use the map as a widget on external websites, applications and such. The only interactive element within the map itself shows when the mouse cursor is moved above a municipality. By doing this additional information is shown on social welfare recipients within that area (see fig. 42).
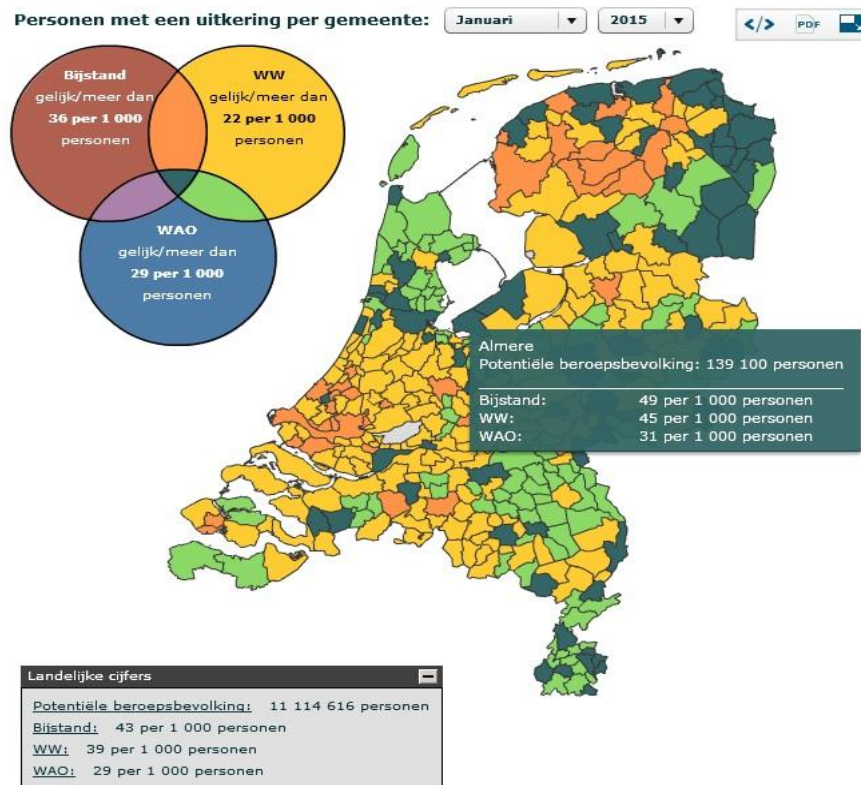
Figure 42: Uitkeringsmonitor Venn Diagram and Welfare Map With Information

Information on the working population and numbers on social benefits/welfare recipients within that area are shown on a scale of one per thousand inhabitants (see fig. 42). This is the only contextual information the user gets to see within the interface. Additionally a legend is presented below the map concerning national figures on the population of the Netherlands. These are the only interactive features within the interface, as no further information is given on the data sources, usage, representation, indicators and/or classifications, and no disclaimer or Terms of Service can be found.

2.3.3 Governance

The governance consists of how the Uitkeringsmonitor seeks to manage and regulate user activity in order to sustain its operating model and fulfil its vision (Light et al. 2016). It can expand from managing user activity to enforcing certain behaviour. It is often enacted through the Terms of Service (ToS) and possibilities the user is given to manoeuvre through the interface, what has been shown in the previous part. The governance is most explicitly described in the disclaimer and ToS accompanying the application, as no ToS and disclaimer are present and accompanying the application, the focus will be held on how governance is enacted through the enforcement of user behaviour and the construction of the interface.

The user only has limited options within the interface, which consists of filling the Venn diagram, selecting a specific colour which is attributed to the selected category of social welfare, and

a measure of one per selected number of inhabitants. Furthermore the user can click on a municipality showing some contextual information, and finally selecting a time frame, full-screen mode or downloading the visualized numbers in a PDF file. This shows that the application is a basic and simple constructed application with not much options regarded the freedom of manoeuvring for the user. The application only visualizes the selected criteria when all three circles are filled in by the user, this means that the interface forces the user to select three social welfare types, or else it will now show the user anything. Finally the choice for a Venn diagram is strange as there are six different social welfare categories, while only three can be selected within the visualization. This means that the user can never get a total image of all social welfare categories within the map. As such the application gives a oversimplified representation of the data on social welfare/benefits recipients, and stays foremost a black-boxed system as it does not explicitly refer to official documentation or is accompanied by a disclaimer, ToS or other contextual information on the workings and functions within the application.

2.3.4 Operational Model

The operational model examines the business strategy and revenue sources of the Uitkeringsmonitor. As no explicit reference is made regarding official documentation, the operational model is somewhat black-boxed. Regarding the implicit reference towards numbers on social welfare/benefits recipients within the Netherlands on the open data Statline platform, and the document 'Arbeidsmarkt in Cijfers' (CBS 2016a), no information can be found on the business strategy or revenue sources of the Uitkeringsmonitor. Furthermore, no tracker cookies are installed when visiting the application, which means that no data is collected on the users.

2.3.5 Technical features:

Software: Adobe Flash Player

Version: 30.0.0.113 (64 bit)

Supported devices: portables (laptop, macbook, ipad, kindle, mobile phones); pc, imacs

Updated: last update in 10-2015

2.4 The BenIkGemiddeld

2.4.1 Scenario of Use

When the user enters the application a text is shown with the following question:"ben jij uniek of ben jij gemiddeld?" This is translated as "are you unique or are you average?" (by author), this is shown in fig. 43.

Figure 43: Ben Ik Gemiddeld Homepage

It is unknown what the user has to do in order to move through the interface, as no buttons or menu is shown. By clicking anywhere on the text the user gets to see the interface, but this is not explicitly mentioned within the application. When the user clicks on the text the following screen is shown (fig. 44).

Figure 44: Ben Ik Gemiddeld Persona Selection

The user has to choose a persona which is illustrated as a cartoon figure, varying in gender, clothes, ethnicity, and style. The user has to choose one of the visualized personas and is not able to create its own persona. The text "wie wil je zijn" is shown which is translated as "who do you want to be". This connotes a different meaning than the question "who are you?", as such the application facilitates a role-playing context in which the user can experiment with different personas to see their average numbers and check if its personas belong the norms. If the user clicks on the persona the following screen is shown (fig. 45).
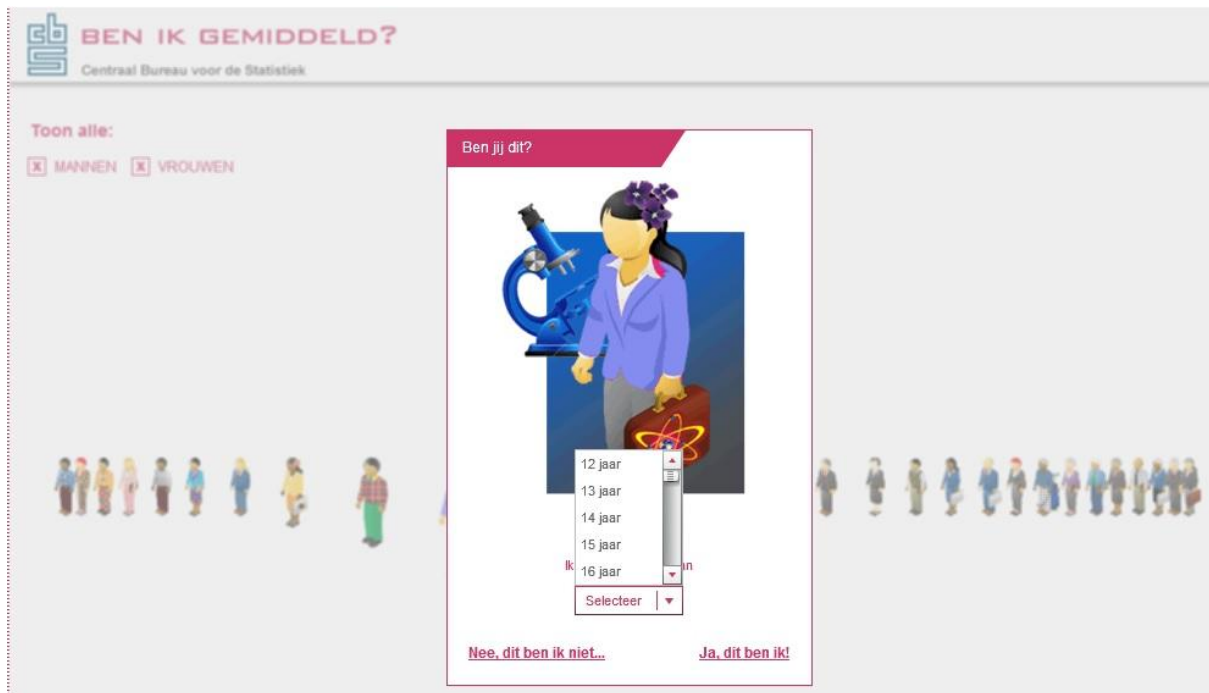
Figure 45: Ben Ik Gemiddeld Persona Specification Window

The question is asked "ben jij dit", which asks if you are this persona. You can select an age ranging from twelve to ninety-nine years old. Furthermore, the job of the persona is represented with a tool or symbol, and with the clothes the persona wears. In this instance this instrument is a microscope, and the persona is nicely dressed with a suitcase with a atom sticker on it. This connotes that the persona is a scholar or active within the scientific community. Then the user has two options: clicking on the button "nee dit ben ik niet" (no this is not who I am); or "ja dit ben ik" (yes this am I). When the user clicks on the former button, the user returns to the persona screen as shown in fig ... When the user clicks on the latter button, the user goes to the next screen within the interface (see fig. 46).

Figure 46: Ben Ik Gemiddeld Theme Selection

The user has to select a theme of which the averages will be shown. The range of selection consist of five dimensions: culture and leisure, population, expenditures, health, and income (translated by author). Within these dimensions a sub-division is made in which the user can further personalize the variables. The application then calculates and visualizes a national average based on the selected criteria. Based on these criteria the user can test if it conforms to the national average (see fig. 47)
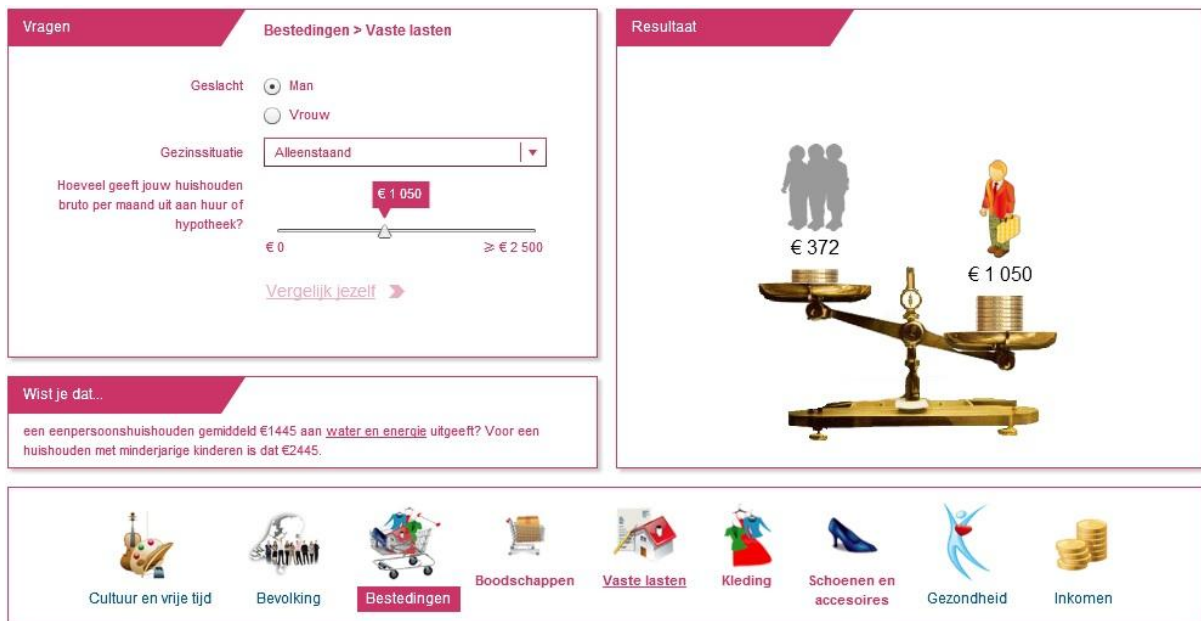
Figure 47: Ben Ik Gemiddeld Interface With Visualization

The user has to answer questions regarding five dimensions: culture and leisure, population, expenditures, health, and income (translated by author). When the answers are given and a specific demographic group is selected based on gender, age and income, the user gets to see visualizations in the form of a chart, bar or a weight scale on which is illustrated if the user belongs to an average demographics group or not. Furthermore depending on the chosen theme, an additional question has to be answered, as seen in fig. 47 the theme "expenses" (translated by author) is chosen. Here an additional question regarding your family status has to be answered, together with a bar on which the user has to answer how much its household spends each month on rent or loans. After answering the user can compare itself to the national average regarding the provided answers. Furthermore, a piece of information is given regarding the selected theme. It is unknown how the personal calculations are constructed, or how the national averages are provided.

The only options the user has is to click on the other themes at the bottom of the screen. Clicking on these themes show four different kinds of visualizations. The one shown in fig. 47, and the ones shown below:

Figure 48: Ben Ik Gemiddeld Interface With Bar Charts

Bar graphs are shown when the dimension Culture and leisure is selected, with your persona represented within one of the bars. The same visualization is used within the population dimension. Within the health dimension two different kinds of data visualizations are used. The one shown here:
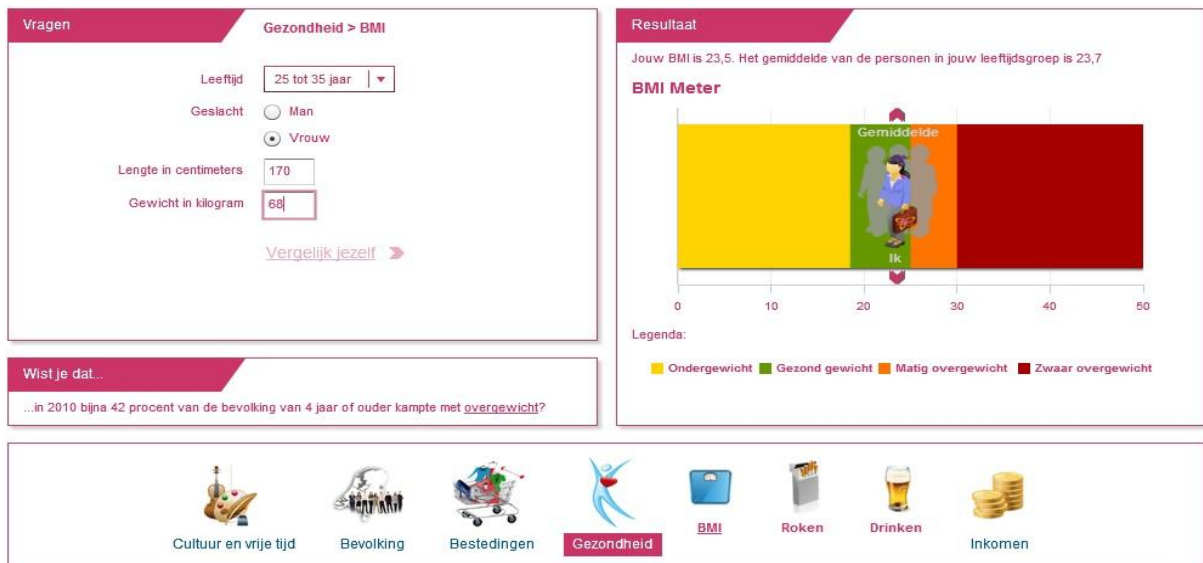

Figure 49: Ben Ik Gemiddeld Interface With Bar Charts Representing BMI

This visualization depicts a bar chart featuring a BMI meter, in where the persona is situated with its selected criteria. Furthermore when clicking on the "roken" or "drinken" buttons the following visualization is shown:
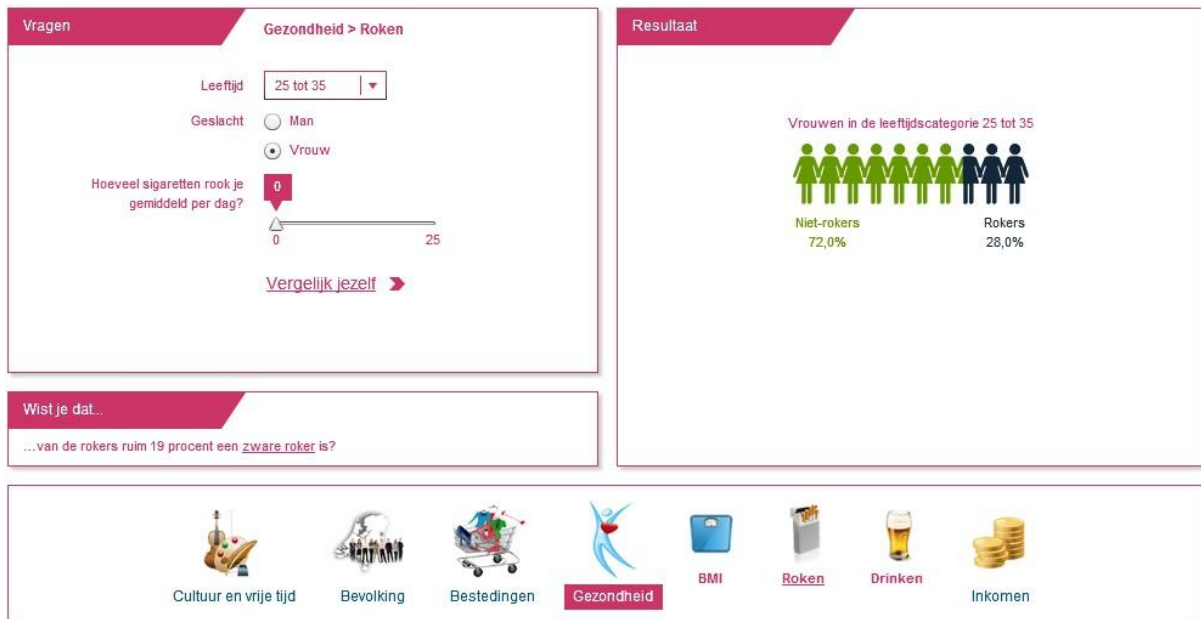
Figure 50: Ben Ik Gemiddeld Interface With People Figures

Here ten small figures are shown which each represent an amount of 10% of the total population sorted by the specified criteria. The green figures show the population that does not smoke, the dark figures show the population which does smoke. Clicking on the income dimension shows the same visualization as depicted in fig. 45 with a weight scale. The only interactive options the user has within each window is the possibility to select an age group, gender, a scale on which it can depict an amount of something, and a button where the user can compare itself to the national average.

2.4.3 Governance

The governance consists of how the BIG application seeks to manage and regulate user activity in order to sustain its operating model and fulfil its vision (Light et al. 2016). It can expand from managing user activity to enforcing certain behaviour. It is often enacted through the Terms of Service (ToS) and possibilities the user is given to manoeuvre through the interface, what has been shown in the previous part. The governance is most explicitly described in the disclaimer and ToS accompanying the application, as no ToS and disclaimer are present and accompanying the application, the focus will be held on how governance is enacted through the enforcement of user behaviour and the construction of the interface.

The user only has limited options within the interface, which consists of selecting a persona, a gender, age group and clicking on various themes which visualize data concerning national figures of those themes. This shows that the application is a basic and simple constructed application with not much options regarded the freedom of manoeuvring for the user. The application only visualizes the selected criteria within a small graph with bar charts or weight scales, this means that the

interface forces the user to select a pre-constructed persona to enact the application. As such the application gives a oversimplified representation of the data on the national average within the five dimensions, and stays foremost a black-boxed system as it does not explicitly refer to official documentation or is accompanied by a disclaimer, ToS or other contextual information on the workings and functions within the application.

2.4.4 Operational Model

The operational model examines the business strategy and revenue sources of the BIG application. As no explicit reference is made regarding official documentation, the operational model is somewhat black-boxed. Regarding the implicit reference towards numbers on social national averages within the Netherlands on the open data Statline platform, no information can be found on the business strategy or revenue sources of the application. Furthermore, no tracker cookies are installed when visiting the application, which means that no data is collected on the users.

2.4.5 Technical features:

Software: Adobe Flash Player

Version: 30.0.0.113 (64 bit)

Supported devices: portables (laptop, macbook, ipad, kindle, mobile phones); pc, imacs

Updated: last update in 11-2015

3. List of visualization applications CBS

Visualisatieoverzicht

- Huizenmarkt
- Ben ik gemiddeld?
- Branchemonitor
- Personen met een uitkering per gemeente
- Duurzame ontwikkeling
- Domeinen van duurzaamheid
- Groene groei
- Prijzen toen en nu
- Belastingopbrengsten
- Spanningsmeter arbeidsmarkt
- Doodsoorzaakverklaring
- Inkomen en bestedingen
- Economische Indicatoren EU
- Geluksmeter

4. Interview Roderik Ponds Leefbaarometer Statistician (in Dutch)
Report Interview:

Wanneer: 09-03-2018

Wie: Roderik Ponds (Atlas voor Gemeenten Utrecht)

Wat: Roderik Ponds is als statistisch expert werkzaam bij Atlas voor gemeenten in Utrecht, en heeft actief meegewerkt aan de ontwikkeling van de Leefbaarometer. Hij is vooral betrokken geweest bij de modelschattingen die aan de basis liggen van een deel van de selectie, en het gewicht van de indicatoren in de uiteindelijke leefbaarheidscores. Daarnaast heeft hij meegewerkt aan deel van de rapportages die uitleg geven over de Leefbaarometer.

Het viel op dat Roderik de Leefbaarometer beschreef als een "verkennend instrument" voor gemeentes om de leefbaarheid te kunnen meten, terwijl deze beschrijving niet is terug te lezen in documentatie relaterend aan de Leefbaarometer. In deze documentatie is te lezen dat de Leefbaarometer gezien kan worden als een "objectief" middel om de leefbaarheid binnen bepaalde gebieden te toetsen. Verder zei Roderik dat de Leefbaarometer specifiek voor gemeentes is ontwikkeld, terwijl in de documentatie is te lezen dat het wel in opdracht van gemeentes is ontwikkeld, maar voornamelijk om een breder publiek inzicht te verschaffen in openbare data van het CBS. Roderik had geen inzicht in cijfers over wie de Leefbaarometer raadplegen. Toen Gerwin en ik dieper ingingen op de politieke keuzes en assumpties die in het systeem zijn gemaakt, relaterend aan de weging van groepen mensen met een bepaalde afkomst in de berekening van de leefbaarheid, was het merkbaar dat Roderik niet eerder op deze manier over de Leefbaarometer had nagedacht. Het was interessant om te zien dat Roderik overtuigd was van de statistische validiteit van de Leefbaarometer, terwijl hij merkbaar verrast was als het ging over de ethische/politieke aspecten van de Leefbaarometer. Roderik was hier echter open over, en vond het dan ook interessant om over deze aspecten na te denken. Roderik vindt de Leefbaarometer een "eerlijk en open" instrument om de leefbaarheid te toetsen, en hij zei dat zij er zoveel mogelijk aan hebben gedaan om de leefbaarheid binnen Nederland op een zo gegrond mogelijke manier te representeren, aan de hand van de beschikbare data. Hij is zich er van bewust dat het altijd beter zou kunnen, en dat bepaalde groepen mensen zich beter gerepresenteerd voelen in de berekening van de leefbaarheid, maar dat dit voornamelijk komt doordat bepaalde data wel/niet beschikbaar is. Het is volgens hem dus onoverkomelijk dat niet de gehele Nederlandse bevolking evenwichtig en "neutraal" gerepresenteerd is in de Leefbaarometer, maar dit door middel van simulaties en regressieanalyses wel zo goed mogelijk is gecompenseerd.