

Article

“I Think Quality is More Important Than a Lot of Data” in Cities Datafication

Carl Chineme Okafor

Department of Media and Social Sciences, University of Stavanger, Norway; carl.c.okafor@uis.no

Submitted: 18 November 2022 | Accepted: 14 June 2023 | Published: 28 June 2023

Abstract

This article studies how the decision to connect data volumes to value is made by technologists and governance people in smart cities’ datafication process. Its entry point is that datafication promises to use data to make cities liveable domains. Cities on the back of this promise presuppose that more data produce value and therefore fixate on exhaustive datafication. But datafication does not appear self-evident, and knowledge of how technologists and governance people connect data volumes to data value is quite unclear in media and communication literature. Using evidence from interviews ($n = 6$), datafication policy documents ($n = 4$), and a diverse dataset of city activities ($n = 299$) in the open data portal of a situated datafication site, the Stavanger Smart City, Norway, and with the theoretical support of critical data studies, this article responds to the question: How does data volume connect to data value in smart cities datafication? Its findings put data quality as the intermediary that makes this connection.

Keywords

data quality; data value; data volumes; smart city datafication; Stavanger Smart City

Issue

This article is part of the issue “A Datafied Society: Data Power, Infrastructures, and Regulations” edited by Raul Ferrer-Conill (University of Stavanger / Karlstad University), Helle Sjøvaag (University of Stavanger), and Ragnhild Kr. Olsen (Oslo Metropolitan University).

© 2023 by the author(s); licensee Cogitatio Press (Lisbon, Portugal). This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

1. Introduction

Exhaustive datafication ensures that cities’ data repositories possess and express knowledge, capital, and power (Eubanks, 2017; Kitchin, 2013, 2021; Ma et al., 2020; Rose, 2020); and this prompts the interrogation of datafication by a critical group of media and communications scholars (e.g., Cardullo & Kitchin, 2019; Zuboff, 2019, pp. 8–14), believing that it is a capitalistic form of control and rationalisation of human mobility (Kitchin, 2021; Sadowski, 2019; Sadowski & Pasquale, 2015).

The crux of the debate is datafication’s promise of a better outcome for modern cities using data, and cities’ huge collection and valuation of data for governance decisions (Beaulieu & Leonelli, 2022, p. 6). Nowadays, cities datafy opinion formation, political debates, and distribution of services and opportunities (Lycett, 2013; Mejias & Couldry, 2019), making them collectors of vast volumes of data. Data is thus the mechanism to opera-

tionalise cities’ communication technologies, but also to redefine engagements therein.

Data indeed possess value, but this article advances that volumes alone do not generate value or sustain cities’ appetite for exhaustive datafication. Often, media and communications scholars focus on the value that volumes of data generate to cities in datafication (Mechant & Walravens, 2018; van der Graaf, 2018) but not specifically on how the connection between volume and value is made by the people who decide on data collection and valuation.

Thus, this article investigates cities’ data collection appetite which is buoyed by the belief that more data will make them better places to live or more efficient to govern. Concretely, it examines if vast volumes of data alone are sufficient to create such value for cities, positing that data needs to be of a certain quality for precise analysis to generate value. By mapping data volumes and data quality, this article sheds light on the potential of

connecting data volumes and data quality in ways that create value.

To understand how value accrues from volumes and cities' propensity for exhaustive datafication, this article studies how decisions on data volumes and value are made by technologists and governance bodies who prioritise data collection and valuation in cities. The main research question in this regard is: How does data volume connect to data value in smart cities' datafication? Findings indicate that data quality enables this connection, contributing to our understanding of the connection between cities' intensive datafication and their appetite for data volume.

2. Literature Review

Smart cities' datafication is popular among governments and technologists (Karvonen et al., 2018; Kitchin, 2016), and scholars try to define smart cities with no common definition yet (Zhao et al., 2021). However, ideas about what a smart city is, can be, and/or should do are not in shortage (Csukás & Szabó, 2021). In this sense, smart cities are conceived, designed, and implemented by governments and technologists with a focus on city efficiency, allowing for larger control on mobility and use of resources, but also generating more data and providing an image of modernity (Al Nuaimi et al., 2015; Hashem et al., 2016)

Datafication is the foundation of smart cities, providing the raw material—data—upon which smart cities initiatives operate (Cardullo & Kitchin, 2019; Kummitha & Crutzen, 2017; Löfgren & Webster, 2020). Datafication claims to make cities smart when technical objects which are previously considered lifeless become cognitively conscious (Akhilesh, 2020). It is a longstanding quantification practice supported by digitisation (Lycett, 2013; Mayer-Schönberger & Cukier, 2013, pp. 74–96; Meijas & Couldry, 2019). Its use in smart cities is extensive.

To “datafy” is to place phenomena into quantified formats, tabulate and analyse for decision-making objectives. van Dijck (2014, p. 198) explains that datafication is based on “a widespread belief in the objective quantification and potential tracking of all kinds of human behavior and sociality through online media technologies.” In other words, everything that can be measured—relationships, experiences, moods—will be turned into a data point and added to a dataset (Ruckenstein & Schüll, 2017, p. 262). But datafication is also met with criticism from scholars (e.g., Chan et al., 2022; Lycett, 2013; Sadowski, 2019; Sadowski & Pasquale, 2015; Zuboff, 2019) who claim that it carries intrusive properties, including increasing citizen surveillance and nudging actions.

The data infrastructures of smart cities rely heavily on digital communication technologies and systems that enable datafication (Mohanty et al., 2016; Rose, 2020). This embeds smart cities' datafication to questions related to the role of communication infrastruc-

tures in modern societies, as they constitute “computing and network resources that allow multiple stakeholders to orchestrate their services and content needs” (Constantinides et al., 2018, p. 381). Datafication practices become intrinsic to decision-making, and thus to power (Sjøvaag & Ferrer-Conill, 2023). To that end, this article adopts Harrison et al. (2010, p. 2) definition of the smart city as “connecting the physical infrastructure, the IT infrastructure, the social infrastructure, and the business infrastructure to leverage the collective intelligence of the city.” This conceptualisation underlines the constitutive description of the so-called modern city as a convener of economies of scale and facilitator of agglomeration. The definition also approaches the smart city as a socio-technical construct (cf. Edwards et al., 2007) which understands humans to inscribe biases in technology and how these biases replicate themselves in the organisation of cities, thus questioning the notion of “smart” but also bringing into dialogue how data is generated and interpreted for use, i.e., data volumes, value, and provenance.

2.1. Data Volumes

Access to data is important to smart cities (Möller & Von Rimscha, 2017); the velocity, volume, and variety of data collection demonstrate this, but also the appetite for data (Beaulieu & Leonelli, 2022, pp. 6–10; Lycett, 2013). This article argues that volume is a derivative of velocity and variety, and contextually connects velocity, volume, and variety to advance that the voracious and varied data cities collect make up data volumes. Volumes in cities connote the size of data that is collected from all available sources for decision-making (Al Nuaimi et al., 2015). Scholars (e.g., van Dijck, 2014; Zuboff, 2019), in their criticism of cities' conviction in governance through aggregated data, highlight that data volumes in cities' datafication process are often seen in the assumption of a manifest relationship between data and efficient cities. In this regard, dialogues on data volumes are often not without thoughts on how and where data comes from, i.e., data provenance, which is important to technologists and governance people in cities' datafication process (Beaulieu & Leonelli, 2022, pp. 7–8).

2.2. Data Value

Data value is the possible advantage cities generate from data volume through analyses and is often projected as the end goal of datafication (Al Nuaimi et al., 2015; Beaulieu & Leonelli, 2022, pp. 24, 39). Value is recognised in how cities mobilise data across contexts, space, and time, and its traces are evident in the technocratic, algorithmic, automated, and anticipatory behaviours cities exhibit (Al Nuaimi et al., 2015; Bibri, 2021; Löfgren & Webster, 2020). Scholars infer that cities recognise data to possess value both in the present and future (Bibri, 2021; Lycett, 2013; Sadowski, 2019; Sadowski & Pasquale, 2015; Rijshouwer

et al., 2022; van Dijck, 2014). In particular, the trust in algorithms to predict and pattern events in cities, but also to automate decisions, demonstrates this inference (Mayer-Schönberger & Cukier, 2013; O’Neil, 2016). Furthermore, cities’ reliance on data to produce and evaluate social knowledge—e.g., “how to parent, police, govern, be healthy and put together a good soccer team” (Beaulieu & Leonelli, 2022, pp. 13–14)—or reinvent governance, are examples of the framings of data value in datafication studies (Kummittha & Crutzen, 2017; Noveck, 2018, pp. 123–126). Likewise, when cities consider data as economic, political, and social means, they, in other words, demonstrate the agency as value that data gain to, for instance, nudge, steer, and control behavioural boundaries or patterns therein (Beaulieu & Leonelli, 2022b; Cardullo & Kitchin, 2019; Kitchin, 2021; Mejias & Couldry, 2019; O’Neil, 2016, p. 191; Sadowski, 2019).

2.3. Data Provenance

In critical media and communication studies on datafication, data volume, and value are expansively discussed as important subjects, but not data provenance. Data provenance defines as the origin of data. It is also described as the conditions under which data is generated and disseminated (Beaulieu & Leonelli, 2022, p. 23). Beaulieu and Leonelli (2022) claim that focusing only on data volumes, without accounting for other important elements (e.g., data provenance) could result in risky data analyses and interpretation. Their contention underscores the need to account for and recognise what happens to data between acquisition and use in generating datafication outcomes.

Scholars in scientific knowledge production often question data collection methods, mostly to account for biases, which involve the properties of data. Similarly, when cities fail to account for the origin of the data they engage with, they risk flawed analysis and claims in decision-making. Data provenance shows the circumstances of data generation to support data use. Like volume and value, it should be a considerable subject in the interrogation of cities’ datafication (boyd & Crawford, 2012). Studying the provenance of data helps to respond to questions of why, how, where, when, and whom of datafication. It supports technologists and governance groups with insights into making decisions on volumes and value. Connecting with volumes and value, the provenance of data articulates the constitutive character of cities’ data assemblage which, as Kitchin and Lauriault (2014) claim, is rooted in conventions, traditions, and infrastructures.

To this end, this article focuses on analysing Data Volumes, Value, and Provenance, in response to its main research question which further opens out to:

- RQ1: What do the technical and governance people in the Stavanger Smart City want from their datafication practice and process?

- RQ2: What is important to the technical and governance people in the Stavanger Smart City datafication practice and process?
- RQ3: How do the technical and governance people in the Stavanger Smart City connect data volume and value to complete the datafication process?

3. Theoretical Framework

Cities’ excitement about big data parallels that of the emergence of statistics in the late 18th century, but critical data studies (CDS) invite media and communications scholars to pay deeper attention to this exciting description of data and the cultures around it (Dalton et al., 2016). Based on datafication’s promises of a better outcome for cities—a claim that influences its ubiquity—this invitation is to interrogate the datafication process beyond how and what cities use data for, to data collection and analysis.

Datafication provides data to smart cities as the raw material to operate, but also as the mechanism to understand their sociality (Dalton et al., 2016; van Dijck, 2014), hence CDS’s call for a critical interrogation of data, their generation, and analysis. Scholars of diverse fields, including media and communications, engage with the assumptions of CDS to respond to the technical and organisational issues that data-intensive practices generate. Similarly, CDS substantially attends to the normative and privacy concerns that cities’ datafication process generates, providing scholars with the theoretical tools to study the widespread consequences of big data in the social arena. From the point of view of data production and analysis, Dalton and Thatcher (2014), for instance, probe the manipulation of big data, including the motives and imperatives that often drive such data work. Other scholars have equally continued to count on CDS to interrogate how cities relate to data volumes and value, but bringing into the conversation how multiple relevant influences embed in the cultures of technological infrastructure political orientations, business, and economic plans or agendas to jointly frame datafication (Iliadis & Russo, 2016; Sadowski, 2019).

In this article, I engage the propositions of the CDS to study the socio-technological process of collecting volumes of data and extracting value from data, but also to know how this process manifests in data assemblages which consist of data systems of technological, political, social, and economic arrangements (Kitchin & Lauriault, 2014). Clarifying how data is consciously and unconsciously created, the means of collecting data, and what informs their analysis in smart cities datafication offers this article the space to bring attention to the entire process of interpreting data for value and how this is decided (boyd & Crawford, 2012).

Of interest to this article is the process of “cooking” data into context-dependent decisions. The “cooking” of data is a key theoretical assumption of CDS and is regarded in this article as the analytical process

of turning data into value. It highlights the misleading idea that data is neutral (Beer, 2016; Gitelman, 2013, pp. 167–171; Räsänen & Nyce, 2013), when they are not because they are often rooted in values, norms, epistemological claims, and philosophical outlooks (boyd & Crawford, 2012) which technologists and governance people in smart cities possess. In other words, all data are produced and, in the process, subject to choices about what to collect and how to analyse them for value (Kitchin, 2014; Sadowski & Bendor, 2019).

Therefore, I operationalise this assumption in studying the connection between data volume (which contextually embodies vast data collection) and value. This way, I investigate the belief that collecting more data is always better for cities to function. More concretely, in engaging the process of “cooking” data, I focus on how technologists and governance bodies who make decisions to aggregate data decide on their values; I also highlight the lack of clarity in scholarship and the social arena about how decisions on data collection and valuation are made (Andrejevic, 2014). This approach underlines the potentiality of datafication to mutate with new ideas and knowledge in cities, for example, when technologies are (re)invented, organisations change, business models are created and recreated, and political systems and economies get altered by new or old orders (Kitchin & Lauriault, 2014).

4. Data and Method

This article adopts a case study approach (Yin, 2018), using the Stavanger Smart City as a situated case. Stavanger is a purposive choice because it is diversely connected with the University of Stavanger as a research hub and provides this research access that would not be possible anywhere else. Stavanger is located in the southwest of Norway; it is both its fourth largest city with approximately 250,000 inhabitants and an energy hub. Stavanger has a smart city operation that dates back to 2016, with goals structured in five priority areas: health and welfare; education and knowledge; energy, climate, and environment; urban art; and governance and democracy (Stavanger City Council, 2016). Stavanger has also achieved smart city goals that include operating an open data portal, weed-control sensors, auto-measuring waste accumulation and disposal, digitising public services fault reporting, and managing city mobility. It has benefitted from the EU-funded programmes Horizon 2020 (2015–2020) for smart cities and communities in Europe and AI4Cities to support climate change goals with artificial intelligence.

Case studies are useful because they respond to the “how” or “why” questions of research. They also come in handy when researchers have limited control over events and the object of study is a contemporary phenomenon with complexities and contextual conditions that need to be studied closely and robustly. In addition, case studies often provide researchers with robust methodological

tools to investigate cases of “decisions,” “individuals,” and “processes” (Yin, 2018). The case study method suits this research as it seeks to understand how technologists and governance people make decisions on data volume and value in a city’s datafication process.

The data collection methods employed in the case study include interviews (Edwards & Holland, 2013), document analysis (Bowen, 2009; Dalglish et al., 2021; Grant & Kara, 2022), and a content analysis of an open data portal (Krippendorff, 2018, pp. 89–124). The methods give access to policy decisions that guide datafication, technologists, and governance people who implement datafication in Stavanger.

I conducted six semi-structured interviews with technologists and governance people in August and October 2022 (see Table 1 for informants’ descriptions). The informants were chosen using maximum variation sampling to get a wide range of interactions from two distinct profiles, experts and practitioners who implement datafication in the smart city (Sandelowski, 1995, 2000). They were recruited from two relevant sources—the Nordic Edge, a smart city and Internet of Things industry cluster associated with Stavanger, with members as vendors and participants in the smart city project, and the Stavanger Smart City Department, which is chiefly responsible for stakeholders’ engagements and evaluation of datafication implementation in the case. These include face-to-face interviews with five participants and one online session on Microsoft Teams with a respondent who was physically unavailable. The interviews were conducted freely in the English language, lasting an average of 38 minutes per session. Respondents’ privacy and ethical use of interview data were guaranteed, as required by the Norwegian Centre for Research Data (Norsk Sentre for Forskningsdata). Consent for this was fully obtained. Interviews covered demographic (name, age, professional status, and work affiliations) and research-related questions. The responses were comprehensive and a 50:50 male and female gender parity was observed.

I transcribed and used the thematic analyses (Braun & Clarke, 2006) to organise themes around my three analytical groupings of Data Volume, Value, and Provenance. The flexibility of thematic analysis allows me to informally determine themes’ prevalence in the interview data, i.e., the prevalence of themes came from repeated data analyses. The theoretical freedom that thematic analysis enjoys also gives it the flexibility to richly account for my data as it reflects and explains the reality and the surface of my data, i.e., making its character transparent. This strategy also captures important themes from the data, with the “keyness” of themes not necessarily depending on quantifiable measures but on relevance to my research questions.

Using the thematic analysis as well, I analysed the four datafication policy papers: (a) the *ICT Strategy for Stavanger Municipality*, (b) the *Joint Social Element of the Municipal Masterplan for New Stavanger 2020–2034* (both were translated from Norwegian to English), (c) the

Table 1. Description of the study’s sample.

Study sample: Interview respondents (IRs) and datafication documents (DDs)	Description
IR01	Governance/technical actor: Data and network engineer, with more than 20 years of experience in digitalisation and innovation; huge involvement in datafication decisions and implementation in Stavanger
IR02	Technical actor: Geodetic engineer, with more than eight years of experience in specialised spatial data management
IR03	Technical actor: Social scientist and IT expert, with more than 10 years of experience in spatial mapping research and innovation
IR04	Governance/technical actor: Technologist, with expertise in cyber security, open data, micro-mobility, citizen involvement, and sustainability in smart cities; more than four years of experience with Stavanger Smart City
IR05	Technical actor: Expertise in smart city systems and applications with more than 10 years of experience in public sensor systems management
IR06	Governance actor: IT innovation strategists with specialisation in digital business, service design, and innovation; more than six years of experience with Stavanger Smart City
DD01	<i>ICT strategy for Stavanger Municipality</i>
DD02	<i>Joint Social Element of the Municipal Masterplan for New Stavanger 2020–2034</i>
DD03	<i>Stavanger Digital Strategy 2014–2029</i>
DD04	<i>Roadmap for Stavanger Smart City</i>

Stavanger Digital Strategy 2014–2029, and (d) the Roadmap for the Smart City Stavanger, to organise themes into my three analytical groups of Data Volume, Value, and Provenance. I used the READ approach: (a) ready your materials, (b) extract data, (c) analyse data, and (d) distil your findings (Dalglish et al., 2021) to do this and gained familiarity with the manifest contents of Stavanger’s datafication plan, including how technologists and governance people understand the plan. I did a content analysis of Stavanger’s open data portal to gain insight into the data that is collected (see Table 2) and identify from its metadata themes that fall into my analytical groupings. In organising my analysis of the data, I relied on the theoretical views of the CDS, specifically observing the “cooking” of data in the process.

5. Results

Two data structures, the open data portal (for the public) and data lake (for internal operations), illustrate Stavanger’s data use practice. However, I did a content analysis of only the open data portal ($n = 299$) which is available to the public. First, I find that eight categories of datasets including an Expired category ($n = 94$) are specific to Stavanger and that there are datasets in duplicates and unusual groupings in contents. I then make new groupings of four (see Table 2) from their original groupings excluding the expired datasets, by uniting and or renaming three categories, i.e., merging Bicycle measurements ($n = 43$) and Transport ($n = 31$) datasets to form Transport and Mobility ($n = 74$); Livelihoods ($n = 25$),

Table 2. Analysis of open data.

Transport and Mobility ($n = 74$)	Maps, Emergency, and Public Safety ($n = 51$)	Culture and Livelihoods ($n = 68$)	Weather and Environment ($n = 12$)
For example, parking, city bike counter, and cycling/hiking routes	For example, mapping, road noise levels, pedestrian traffic, school routes, vehicle charging stations, and speed limits	For example, registered unemployment/disability pensioners, immigration, and operating expenses for museums and cinemas	For example, waste management/refuse containers, rain/manhole/bathing temperature sensors, and air quality measurement

Culture ($n = 12$), and Statistics ($n = 31$) to form Culture and Livelihoods ($n = 68$); and renaming Maps, Emergency, and Public Safety ($n = 51$). Weather and Environment ($n = 12$) retain their categorisation. This was to make my analysis transparent and understandable.

Excluding the 94 expired datasets, the portal contains 207 datasets that are actively specific to Stavanger. Their metadata provides insights on the dates of collection (starting from 2016), update schedules (every two and five minutes for parking and city bikes data respectively, hourly for air quality data, daily for cycling and rain data, annually and on-demand for hiking routes and trails data), the protocol that governs data use and reuse (the Norwegian Licence for Open Government Data) and data storage formats (CSV, JavaScript Object Notation, GPS Exchange, Microsoft Word text document, and PDF).

5.1. Thematic Analysis

I use the thematic analysis to analyse the six interview transcripts and four datafication policy papers which are part of my empirical materials. The documents are (a) the *ICT strategy for Stavanger Municipality*, (b) the *Joint Social Element of the Municipal Masterplan for New Stavanger 2020–2034* (both were translated from Norwegian to English), (c) the *Stavanger Digital Strategy 2014–2029*, and (d) the *Roadmap for the Smart City Stavanger*. In addition to providing me with historical insights into Stavanger’s datafication goals and implementation context, I find in these materials 10 themes that fall into my three analytical groupings of Data Volume, Value, and Provenance (see Table 3).

Data Volume is discussed through five themes. Data Volume in the context of smart city datafication connotes the size and variety of data collected from all available collection sources. Data from documents and interview analysis describe Stavanger’s data collection practice, but also the number of datasets and the varied formats they are stored in the open data portal. One of the prom-

inent themes concerns managing large amounts of data for society and its citizens. As informant DD01 states:

With control and an overview of the data, it is easier for the municipality to use data in new contexts such as artificial intelligence, data analysis, and big data. More data can be compiled and create new insights and improved services. The municipality will also share data [open data] with other players to contribute to innovation and service improvement for the benefit of citizens. The municipality must therefore have discretion, control, and access to all data that the municipality produces.

Another aspect of Data Volume in the themes concerns how new technologies will continue to produce more data, as informant DD01 again explains:

The municipality manages large amounts of data on behalf of society and its citizens, and new technology will produce even more. The investment in open data [data sharing] will continue so that we facilitate innovation and reuse also outside the municipality.

Also, Volume is the theme that reflects on how data collection is based on legal, economic, and clerical mandates of Stavanger’s seven service departments, and informant IR01 explains this thus:

It is very individualistic, from department to department. It depends on the tasks and services that they have the mandate to operate. For instance, the garbage people [Waste Disposal Unit] have data on different garbage bins. To the point, that would be what kind of clerical mandate we are set under.

In addition, informant IR04 underlines the extent of this theme saying: “When it comes to ownership and maintenance of data, it is up to different departments that have different professional fields, and there will be

Table 3. Themes from analysis of empirical data.

Analytical groups	Themes from datafication policy papers and interview transcripts
Data Volume	<ol style="list-style-type: none"> 1. Municipality manages large amounts of data for society and its citizens 2. New technology will produce more data 3. Data collection is based on legal, economic, and clerical mandates 4. Data architecture and infrastructure must be built for data exchange and communication to have a special focus on scalability, openness, and interoperability 5. Municipality has standard data archiving and reuse process
Data Value	<ol style="list-style-type: none"> 1. Data as an important resource is used for governance 2. The municipality must ensure ownership and access to data to build predictive decision-making 3. Data will be used and reused in other contexts that can provide new and better services in the future
Data Provenance	<ol style="list-style-type: none"> 1. The quality of data collected is important for value generation 2. Quality data is realised from iteration and standardisation of collection

people for owning such data. Although, in theory, the city administrator owns all the data.”

Further in relation to Data Volume is a theme expressing how the city’s data architecture and infrastructure must be built for data exchange and communication and configured towards scalability, openness, and interoperability as supported by the governing protocol data use in the open data portal. On this, DD01 explains:

Solid architecture and well-thought-out infrastructure are pillars of forward-looking, good citizen services, and effective ICT tools for the employees. This is therefore the municipality’s main focus area within ICT. The architecture and infrastructure must be built so that it is possible to exchange data and communicate between different ICT systems, also across administrative levels. New ICT acquisitions, further development, and resource requirements depend on the ICT architecture. It must be facilitated for an efficient, modular, flexible, and service-oriented architecture and have a special focus on scalability, openness, and interoperability.

And finally, Volume is a theme that relates to the city’s standard data archiving and reuse process, which informant DD01 explains:

All systems that store archive-worthy material must have a strategy for handing over electronic material to Stavanger City Archives. Systems that are not approved as electronic archives can be considered integrated directly with the case/archive system. Reuse of master data across subject systems greatly contributes to saving resources and streamlining work processes, we must therefore decide which system/service is responsible for which data. The information must be registered and maintained in only one place and then made available to all systems that need the information.

In summary, the themes agreeably represent Data Volume and how Stavanger conceives and deploys it in its datafication process.

In terms of Data Value, three themes manifest in the analysis. But this also manifests in the protocol that governs data use and reuse in the open data portal analysis. Data value is the advantage that the smart city stands to generate from data volume. It is usually the end goal of cities’ datafication process. One theme expresses this in the analysis as data being an important resource for governance, and informant DD01 confirms: “Data is an important resource today but will grow both larger and more important in the digital future.”

Informant IRP01 also notes in this same breadth that: “We are collecting the data, putting them together across all the different sources in kind of a dashboard and reports to give us a better governance insight...data that help us back up the goals that we try to achieve.”

Another theme that speaks to Data Value is the municipality’s ownership and access to data to build predictive decision-making. Informant DD01 underscores this: “The municipality must ensure ownership and easy access to this important resource. In the future, the data will be used and reused in other contexts that can provide new and better services.”

Informant IRP01 also explains:

We have a lot of data in the municipality; we have water data, sensor data, and for different solutions—It could be HR, economic data, [and] health department which has a lot of data regarding our citizens. The purpose is to collect all the data and then we have a toolbox that gives us the opportunity to use the data for different purposes. We are starting to use the data for these new technologies on machine learning and prediction.

In agreement, Informant IR04 states: “Most of the operational part of the municipality is data-driven...a lot of the data that is collected within the municipality is to create statistical models for projections.”

Similarly, a theme that talks about Data Value in the analysis is the expectation that data will be used in different contexts and levels to generate competent judgments. Informant DD01 references this: “Through analyses and big data, we can produce good decision-making information at all levels.”

In addition to this, informant DD04 states: “Technology will be part of the solution—Whether new technology is used or existing technology is further developed...technology in a smart city context is a tool for creating economic, social, and environmental improvements.”

Consequently, the themes correspond to the advantages Stavanger expects from the volumes of data it generates.

When it comes to Data Provenance, which is the origin of and conditions under which data is generated and disseminated, two themes describe it in the analysis. The data collection dates and update schedule in the open data portal likewise express Provenance. One of the themes relates to how the quality of data that the city collects is important for value generation. This is as informant IR06 explains:

We don’t need to collect more than we are using...I think that [the] quality of the data is more important than having a lot of data...I think [that] quality data is data that somebody finds useful to make their tasks better at optimising something. But also, that [it] is collected and measured in a good way.

In agreement, informant IR03 explains that:

This might sound like it comes from the school-books and maybe it does; the main objective is

to have FAIR data—findable, accessible, interoperable, and reusable. Quality as such is more related than [the] traditional perspectives on scaling levels and magnitude.

Furthermore, Data Provenance manifests in how iteration and standardisation of data collection result in quality data, and informant IRP01 explains this:

We need at least to standardise the way that we are using data so that they make sense for us. If the systems we use are used in different ways for different scenarios, it will not make any sense to us, [and] if we collect the data because they are not consistent. If we are not standardising the way that we put data in the systems, we don't have good quality data, it loses a lot of quality and it is not good enough to be used. I would say that one of the main struggles in the municipality is that we are not standardised in our work processes, so the quality of the data that we put in will be different and it becomes confusing data.

Informant IR03 shares the same perspective, saying:

This is along the discussion always, and it is quality meaning more detail, more accuracy, and precision. We have another word which might be related to quality—authoritative data—and it is not the same [as quality], but it means that these are data that you can trust, or they come from authority-level processes. For instance, the property data is one of Norway's most known authoritative datasets, and that compared to crowd-sourced or sensor-based data that haven't actually been qualified [to be sure] if they are correct. That means that it is part of the quality sign that the data is authoritative, meaning that it can be trusted, that it has been collected correctly or according to the laws, instructions, or standards that specify the necessary requirements.

Quality as a component for assessing trust and validity in data thus summarises Data Provenance in the analysis. It is also a component of Provenance and indicates that, in Stavanger's datafication process, data quality is functional in connecting volume to value.

6. Discussion

The results presented earlier contribute to my understanding of the datafication process in the Stavanger Smart City, specifically how technologists and governance people make decisions on data volume and value. But before I discuss this further, I would like to follow the scientific virtue of clarifying to highlight how the two actors I engaged with view datafication and the smart city concepts in their everyday operations.

There is no shared definition of the smart city in principle and practice; instead, actors define it accord-

ing to their respective needs. Stavanger has a selective approach to accomplishing its smart city goals, it's understanding of the smart city nevertheless aligns with this article's definition as a city engaging with data and digital communications infrastructure (Harrison et al., 2010, p. 2). While datafication appears to be theoretically defined and settled in scholarship (Mayer-Schönberger & Cukier, 2013; Mejias & Couldry, 2019; Lycett, 2013; van Dijck, 2014, p. 198), I find that this is not the case amongst actors in Stavanger who are unclear about the theoretical label despite their practice of it. This implies that datafication in smart cities is context-driven (Lycett, 2013; Mejias & Couldry, 2019; Micheli et al., 2020).

Following from this, I also consciously point out for clarity that Stavanger's data lake, which is exclusive and unavailable for this study's analysis, however, comprises data from multiple systems and sources that are plugged into the city's digital infrastructure network. Respondents claim that data from therein are internally used to modulate and model perspectives and insights for city management. The data lake and open data portal differ from each other in terms of content, access, and use, but my findings from the open portal analysis show that Stavanger, like most smart cities setup, is compulsively obsessed with data volume, and its value as a governance tool (Al Nuaimi et al., 2015; Ma et al., 2020; Sadowski & Bendor, 2019). The results also indicate that actors have an idea of the importance of data provenance in decisions about data use or engagement, especially as it concerns the open data portal (Beaulieu & Leonelli, 2022, p. 23).

While substantiating the need for this study on the basis of the deficit of empirical thoughts on how decisions to connect volume and value are made in datafication despite extensive literature on datafication, my results establish that Stavanger has a clear datafication program to deliberately collect and generate value from data, trusting it and digital communications infrastructure to drive governance and societal wellbeing (cf. Beaulieu & Leonelli, 2022; Lycett, 2013). In this regard, my analysis suggests that the "cooking" of data is usually required to connect volume to value in Stavanger and that the provenance of data possibly plays an essential role in making this connection (Beaulieu & Leonelli, 2022, pp. 7–8). As an influenced process (Iliadis & Russo, 2016; Sadowski, 2019), datafication in Stavanger involves the "cooking" of data to match goals. In this case, I take note of how respondents talk about data quality—a component of provenance (boyd & Crawford, 2012)—as the intermediary for translating volume into value.

My results show that data is a valuable resource for Stavanger to productively engage, create opportunities (e.g., through its open data portal; cf. Gilbert, 2021), and fix the city's challenges (Kummitha & Crutzen, 2017; Noveck, 2018; van Dijck, 2014; Zuboff, 2019), but the connection between volume and value is achieved through quality, suggesting thus that quality serves an evaluative role in linking value to volume.

In other words, the results imply that value from large volumes of data is accessed through the quality of data that is collected and available to the city but also that the absence of quality could result in no value from volume.

Also based on the results of this study, data quality appears to empower datafication actors in Stavanger to, in the search of value, generate and obsess for more data. This, as respondents imply, is an iterative process to attain data quality, i.e., actors relying on iteration to get quality-level data that support their goal. Simply defined as data that is fit for purpose (Fox et al., 1994), data quality is also labelled by respondents as good data, i.e., data that is of high quality and which can enable good decisions, a claim that further accounts for its grounding in evaluation, but also infers that, in context, it is an ongoing process of improving data for value. This iterative search for value through quality, is in fact, the “cooking” that scholars refer to for data to align with contexts and priority-informed decisions on datafication (Hacking, 2007; Kitchin, 2021; Kitchin & Lauriault, 2014; Löfgren & Webster, 2020; Zuboff, 2019).

7. Conclusion and Contributions

The intensifying dimensions of datafication signify that it is not a passive process but an actively defined practice wherein cities decide on data that affords them specified value. This has become much more sophisticated with time, technology, and agenda and deployed to distribute opportunities, secure societies, and manipulate and modify social actions. In fact, it would appear that its essence is to enable societies to forecast and control their affairs using tons of data and communication infrastructure, a notion that has for long placed data volume and value in front of debates that concerns its premises. In these debates though, less attention has gone into discussing how decisions to connect volume to value are made by active actors of datafication. Having set out to study this—how decisions on data volume and value are made—through two essential actors (technologists and governance people) in a situated datafication site, the Stavanger Smart City, I have in this article made efforts to provide an empirically grounded argument that positions data quality as the intermediary for translating volume into value. My findings do not exclusively make claims that data quality is the only intermediary to translate volume into value but posit that it enables this connection.

While contributing to our understanding of how technologists and governance people make decisions on data volume and value in datafication, I further argue that data quality may as well be instrumental to cities’ appetite for volumes of data.

My results hint at how data quality as a component of data provenance accentuates the role that provenance impliedly plays in establishing the reliability or otherwise of data. I reckon however that the seeming lim-

ited engagement with data quality amongst media and communications studies scholars may have provided the grounds for the many questions that datafication of the social arena generates. These questions are nonetheless vital in contemplation of datafication promises of a better outcome for modern cities, yet scholars’ understanding of data quality as an intermediary to translate volume into value in datafication provides additional knowledge for interrogating datafication, more so leveraging on the theoretical sagacity of CDS.

8. Limitations and Future Research Recommendations

I do not make claims that this study answers all the questions about how the datafication process in cities can be understood, but I have made efforts to study how data quality may contribute to understanding datafication and that scholars can approach future studies of datafication from this perspective. In this breadth, I thus recognise and highlight that there are obvious limitations in this study, one of which is its reliance on a single case study, and particularly the low number of IRs captured in this study. A higher interview sample size ($n = 12$) was planned for, but saturation was achieved midway, yet this shortcoming is compensated with the use of complementary data sources to ground the findings. Future research can concretely investigate the likely dimensions of data quality in cities’ datafication, to establish how the technologists and governance people who decide on data interpret data quality from their different work areas and if they have similar or dissimilar understandings of data quality, as well as how they manage to reconcile potential differences in interpretations of data quality to achieve data interoperability and sharing to meet cities datafication goals.

Acknowledgments

This work was supported by the Norwegian Research Council (Project No. 314257).

Conflict of Interests

The author declares no conflict of interests.

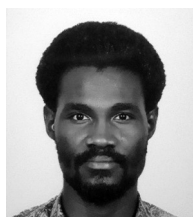
References

- Akhilesh, K. B. (2020). Smart technologies—Scope and applications. In K. B. Akhilesh & D. P. F. Möller (Eds.), *Smart technologies* (pp. 1–16). Springer. https://doi.org/10.1007/978-981-13-7139-4_1
- Al Nuaimi, E., Al Neyadi, H., Mohamed, N., & Al-Jaroodi, J. (2015). Applications of big data to smart cities. *Journal of Internet Services and Applications*, 6(1), Article 25. <https://doi.org/10.1186/s13174-015-0041-5>
- Andrejevic, M. (2014). The big data divide. *International Journal of Communication*, 8(17), 1673–1689.

- Beaulieu, A., & Leonelli, S. (2022). *Data and society: A critical introduction*. SAGE.
- Beer, D. (2016). How should we do the history of big data? *Big Data & Society*, 3(1). <https://doi.org/10.1177/2053951716646135>
- Bibri, S. E. (2021). Data-driven smart sustainable cities of the future: Urban computing and intelligence for strategic, short-term, and joined-up planning. *Computational Urban Science*, 1(1), Article 8. <https://doi.org/10.1007/s43762-021-00008-9>
- Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2), 27–40. <https://doi.org/10.3316/QRJ0902027>
- boyd, d., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5), 662–679. <https://doi.org/10.1080/1369118X.2012.678878>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Cardullo, P., & Kitchin, R. (2019). Smart urbanism and smart citizenship: The neoliberal logic of “citizen-focused” smart cities in Europe. *Environment and Planning C: Politics and Space*, 37(5), 813–830. <https://doi.org/10.1177/0263774X18806508>
- Chan, J., Sanders, C., Bennett Moses, L., & Blackmore, H. (2022). Datafication and the practice of intelligence production. *Big Data & Society*, 9(1). <https://doi.org/10.1177/20539517221089310>
- Constantinides, P., Henfridsson, O., & Parker, G. G. (2018). Introduction—Platforms and infrastructures in the digital age. *Information Systems Research*, 29(2), 381–400. <https://doi.org/10.1287/isre.2018.0794>
- Csukás, M. S., & Szabó, R. Z. (2021). The many faces of the smart city: Differing value propositions in the activity portfolios of nine cities. *Cities*, 112, Article 103116. <https://doi.org/10.1016/j.cities.2021.103116>
- DalGLISH, S. L., Khalid, H., & McMahon, S. A. (2021). Document analysis in health policy research: The READ approach. *Health Policy and Planning*, 35(10), 1424–1431. <https://doi.org/10.1093/heapol/czaa064>
- Dalton, C. M., Taylor, L., & Thatcher, J. (2016). Critical data studies: A dialog on data and space. *Big Data & Society*, 3(1). <https://doi.org/10.1177/2053951716648346>
- Dalton, C., & Thatcher, J. (2014). What does a critical data studies look like, and why do we care? Seven points for a critical approach to “big data”. *Society and Space*, 29.
- Edwards, R., & Holland, J. (2013). *What is qualitative interviewing?* Bloomsbury Academic.
- Edwards, P., Jackson, S., Bowker, G., Knobel, C. (2007). *Understanding infrastructure: Dynamics, tensions, and design*. Santa Clara University.
- Eubanks, V. (2017). *Automating inequality: How high-tech tools profile, police, and punish the poor* (1st ed.). St. Martin’s Press.
- Fox, C., Levitin, A., & Redman, T. (1994). The notion of data and its quality dimensions. *Information Processing & Management*, 30(1), 9–19. [https://doi.org/10.1016/0306-4573\(94\)90020-5](https://doi.org/10.1016/0306-4573(94)90020-5)
- Gilbert, S. (2021). *Good data: Power, paranoia, and prosperity in the digital age*. Welbeck.
- Gitelman, L. (2013). *“Raw data” is an oxymoron*. The MIT Press.
- Grant, A., & Kara, H. (2022). *Doing your research project with documents: A step-by-step guide to take you from start to finish*. Policy Press.
- Hacking, I. (2007). Kinds of people: Moving targets. In *Proceedings of the British Academy Lectures* (Vol. 151, pp. 285–318).
- Harrison, C., Eckman, B., Hamilton, R., Hartswick, P., Kalaganam, J., Paraszczak, J., & Williams, P. (2010). Foundations for smarter cities. *IBM Journal of Research and Development*, 54(4), 1–16. <https://doi.org/10.1147/JRD.2010.2048257>
- Hashem, I., Chang, V., Anuar, N., Adewole, K., Yaqoob, I., Gani, A., Ahmed, E., & Chiroma, H. (2016). The role of big data in smart city. *International Journal of Information Management*, 36(5), 748–758.
- Iliadis, A., & Russo, F. (2016). Critical data studies: An introduction. *Big Data & Society*, 3(2). <https://doi.org/10.1177/2053951716674238>
- Karvonen, A., Cugurullo, F., & Caprotti, F. (Eds.). (2018). *Inside smart cities: Place, politics and urban innovation* (1st ed.). Routledge. <https://doi.org/10.4324/9781351166201>
- Kitchin, R. (2013). Big data and human geography: Opportunities, challenges, and risks. *Dialogues in Human Geography*, 3(3), 262–267.
- Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1). <https://doi.org/10.1177/2053951714528481>
- Kitchin, R. (2016). *Getting smarter about smart cities: Improving data privacy and data security*. Department of the Taoiseach.
- Kitchin, R. (2021). *Data lives: How data are made and shape our world* (1st ed.). Bristol University Press.
- Kitchin, R., & Lauriault, T. P. (2014). *Towards critical data studies: Charting and unpacking data assemblages and their work*. SSRN. <https://ssrn.com/abstract=2474112>
- Krippendorff, K. (2018). *Content analysis: An introduction to its methodology* (4th ed.). SAGE.
- Kummitha, R. K. R., & Crutzen, N. (2017). How do we understand smart cities? An evolutionary perspective. *Cities*, 67, 43–52.
- Löfgren, K., & Webster, C. W. R. (2020). The value of big data in government: The case of “smart cities.” *Big Data & Society*, 7(1). <https://doi.org/10.1177/2053951720912775>
- Lycett, M. (2013). “Datafication”: Making sense of (big)

- data in a complex world. *European Journal of Information Systems*, 22(4), 381–386.
- Ma, M., Preum, S. M., Ahmed, M. Y., Tärneberg, W., Hendawi, A., & Stankovic, J. A. (2020). Data sets, modeling, and decision making in smart cities: A survey. *ACM Transactions on Cyber-Physical Systems*, 4(2), 1–28. <https://doi.org/10.1145/3355283>
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think* (1st ed.). Houghton Mifflin Harcourt.
- Mechant, P., & Walravens, N. (2018). e-Government and smart cities: Theoretical reflections and case studies. *Media and Communication*, 6(4), 119–122. <https://doi.org/10.17645/mac.v6i4.1848>
- Mejias, U. A., & Couldry, N. (2019). Datafication. *Internet Policy Review*, 8(4). <https://doi.org/10.14763/2019.4.1428>
- Micheli, M., Ponti, M., Craglia, M., & Berti Suman, A. (2020). Emerging models of data governance in the age of datafication. *Big Data & Society*, 7(2). <https://doi.org/10.1177/2053951720948087>
- Mohanty, S. P., Choppali, U., & Kougiannos, E. (2016). Everything you wanted to know about smart cities: The Internet of Things is the backbone. *IEEE Consumer Electronics Magazine*, 5(3), 60–70. <https://doi.org/10.1109/MCE.2016.2556879>
- Möller, J., & Von Rimscha, M. B. (2017). (De)centralization of the global informational ecosystem. *Media and Communication*, 5(3), 37–48. <https://doi.org/10.17645/mac.v5i3.1067>
- Noveck, B. S. (2018). Open data: The future of transparency in the age of big data. In D. E. Pozen & M. Schudson (Eds.), *Troubling transparency* (pp. 206–225). Columbia University Press.
- O’Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy* (1st ed.). Crown Books.
- Räsänen, M., & Nyce, J. M. (2013). The raw is cooked: Data in intelligence practice. *Science, Technology, & Human Values*, 38(5), 655–677.
- Rijshouwer, E. A., Leclercq, E. M., & van Zoonen, L. (2022). Public views of the smart city: Towards the construction of a social problem. *Big Data & Society*, 9(1). <https://doi.org/10.1177/20539517211072190>
- Rose, G. (2020). Actually-existing sociality in a smart city: The social as sociological, neoliberal and cybernetic. *City*, 24(3/4), 512–529.
- Ruckenstein, M., & Schüll, N. D. (2017). The datafication of health. *Annual Review of Anthropology*, 46(1), 261–278.
- Sadowski, J. (2019). When data is capital: Datafication, accumulation, and extraction. *Big Data & Society*, 6(1). <https://doi.org/10.1177/2053951718820549>
- Sadowski, J., & Bendor, R. (2019). Selling smartness: Corporate narratives and the smart city as a sociotechnical imaginary. *Science, Technology, & Human Values*, 44(3), 540–563.
- Sadowski, J., & Pasquale, F. (2015). The spectrum of control: A social theory of the smart city. *First Monday*, 20(7). <https://doi.org/10.5210/fm.v20i7.5903>
- Sandelowski, M. (1995). Sample size in qualitative research. *Research in Nursing & Health*, 18(2), 179–183.
- Sandelowski, M. (2000). Whatever happened to qualitative description? *Research in Nursing & Health*, 23(4), 334–340.
- Sjøvaag, H., & Ferrer-Conill, R. (2023, March 14). Datacentre er en del av den digitale infrastrukturen—Og må reguleres deretter [Data centres are part of the digital infrastructure—And must be regulated accordingly]. *Stavanger Aftenblad*. <https://www.aftenbladet.no/meninger/debatt/i/KnbjdM/datasentre-er-en-del-av-den-digitale-infrastrukturen-og-maa-reguleres-deretter>
- Stavanger City Council. (2016). *Roadmap for the Smart City Stavanger. Vision, goals and priority areas*.
- van der Graaf, S. (2018). In Waze we trust: Algorithmic governance of the public sphere. *Media and Communication*, 6(4), 153–162. <https://doi.org/10.17645/mac.v6i4.1710>
- van Dijck, J. (2014). Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveillance & Society*, 12(2), 197–208.
- Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). SAGE.
- Zhao, F., Fashola, O. I., Olarewaju, T. I., & Onwumere, I. (2021). Smart city research: A holistic and state-of-the-art literature review. *Cities*, 119, Article 103406.
- Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.

About the Author



Carl Chineme Okafor is at this time a doctoral candidate in the Department of Media and Social Sciences at the University of Stavanger, Norway. He is a 2019 journalism fellow of the Reuters Institute for the Study of Journalism (RISJ)—the University of Oxford, UK. He has two MA degrees, in international diplomacy, and journalism and social communication. Datafication and journalism are his current research focus.