Datafication in Smart Cities: Understanding How the Public Experience Urban Environments

Jie Qi, Suvodeep Mazumdar, Ana C Vasconcelos
Information School, University of Sheffield, Sheffield, UK

Jqi2@sheffield.ac.uk
s.mazumdar@sheffield.ac.uk
a.c.vasconcelos@sheffield.ac.uk

Abstract: Datafication has become a prominent feature of smart cities, where sensors, monitoring devices, and AI are being integrated with city infrastructures and facilities, resulting in rapidly changing urban areas informed by data-driven decision-making processes. Although there is a vast amount of data being generated about urban environments and citizens, research on understanding citizens' social experience in smart cities has been limited. This study proposes a three-stage research design that provides datafication solutions to understand citizens' experience of urban environment in a synergistic manner. We employ a mixed methods approach drawing upon multiple data collected by the researcher, from the citizens, and across smart cities open data platforms. It is designed to undertake a place-based and citizen-centric approach to understand the lived social experiences of citizens in urban environments. This work will contribute to our current understanding in developing socially sustainable smart cities, providing methodological insights for future research on how datafication process can be leveraged to improve quality of urban life.

Keywords: datafication, open data, online survey, spatial statistical modelling, moderation and mediation analysis, text mining

1. Introduction

The smart city concept has recently become closely associated with quality of life in terms of improving urban liveability (Macke, Casagrande, Sarate, & Silva, 2018), achieving social sustainability (Aurigi & Odendaal, 2021), and fostering a citizen-centric urban development (Andreani, Kalchschmidt, Pinto, & Sayegh, 2019 via knowledge-based solutions. As such, smart cities require publicly accessible datasets (Neves, Neto, & Aparicio, 2020). The datafication of urban life in a smart city consists of a wide range of living experience emerging from city infrastructure, service delivery, and urban amenity. In this regard, it is critical for us to understand how citizens experience and perceive urban environments in the context of social space (Balsa-Barreiro, Menendez, & Morales, 2022), and particularly, urban public spaces. Urban public spaces are the social settings for people to engage in interactions and facilitate 'good urban living' which can effectively (re)vitalise urban areas and are essential for the success of smart cities development (Holland, Clark, Katz, & Peace, 2007). However, despite smart sustainable urbanism becoming increasingly data-driven, there has been very little research effort in datafication pertaining to citizens' experience and the social performance of urban space. Therefore, this paper aims to address the research gap by proposing a mixed research approach that harnesses the strengths of behaviour mapping, open government data, and volunteered geographic information to study user experiences and perceptions in public spaces.

2. Related work

As information technology is embedded in our built environments, the continuous collection of data by deploying digital transformation and technological solutions across sustainable development and management practices aims to make our cities more efficient and liveable. Smart cities, integrating pivotal technologies such as Internet of Things (IoT), cloud computing, big data, and AI, now not only encompass the characteristics of digital city (i.e., datafication of urban practices, technocentric planning solutions, etc.), but are also meant to be conducive to the social and economic transformation required to meet the needs of citizens (Zhao & Zhang, 2020).

While some scholars critique that the first generation of smart city development fell short in exploiting a citizen-centric agenda (Trencher, 2019), many point to a bold ambition to tackle the social challenges brought by the techno-economic and centralised approach of recent smart city interventions (Appio, Lima, & Paroutis, 2019). Some key dimensions related to the social challenges of smart city have been explored in the literature including social inclusion in affordable technology infrastructural services, maintaining social capital in smart neighbourhood management (Nakano & Washizu, 2021), environmental equity (Monfaredzadeh & Krueger, 2015), and social sustainability in engaging with the livelihoods of the marginalised citizens (Aurigi & Odendaal,

2021). Notwithstanding the implications of developing socially progressive and sustainable smart cities, Aurigi & Odendaal (2021) calls for the need of revealing empirical evidence using context-sensitive and place-based approaches that are essential for enhancing the social sustainability of smart cities. Adding to, and complementing these observations, we argue that there is a data-driven and citizen-centric discourse in the smart city paradigm in response to the unique urban built environment forms (spatial context), everyday social interactions (social context) and urban lifestyles. The gaps in exploring datafication to transform domains of urban life and understand social performance of the smart cities urban environment, still prevails at the expense of making smart cities more liveable and socially sustainable.

3. Methodology

This study proposes a three-stage research design that aims to understand citizens' experience of urban environment in a synergistic manner by employing multiple data sources in response to understanding the datafication process in smart cities. The datafication process designed consists a collection of spatial, statistical, and text data that will be digitised and processed using (open sourced) GIS applications and Python, outlines in Table 1, We then discuss how they offer the complementary views on understanding how the public experience urban environment.

Table 1: Detailed description of the proposed mixed methods.

Data Components	Datafication stage 1: **Data Collection**	Datafication stage 2: **Data Preparation**	Datafication stage 3: **Data Processing**
Social Interactions	***Primary spatial data*** collected by quantifying social interaction via a ***behaviour mapping*** exercise.	Digitisation: georeferencing social interaction points using ***ArcGIS***.	*Spatial analysis*: hotspot analysis (spatial autocorrelation analysis).
Urban Public Spaces	1.***Sheffield City Council Open Data*** – *smart cities open data* collected by local authorities; 2.***Ordnance Survey MasterMap*** – *authoritative open data* collected by national mapping agency; 3. ***OpenStreetMap*** – *crowdsourced data / volunteered geographical information (VGI)* collected via online platform.	Spatial Preprocessing in ***QGIS***: 1. geographical feature overlay; 2. interpolation; 3. feature selection analysis.	*Spatial statistical modelling*: 1. buffer analysis; 2. road network analysis; 3. geographically weighted regression analysis.
Citizen Perceptions	***Primary statistical and text data*** representing citizens' perspectives on their social experience in urban environments collected via *online open-ended survey*.	Data preprocessing in ***Python***: 1. feature handling: numeric and categorial value; 2. handling missing value and duplicates; 3. dealing with outliers.	*Moderation and mediation analysis* *Text mining*: sentiment analysis with machine learning (ML) based approach using Python NLTK, an open-sourced library for natural language processing (NLP) in Python.

As shown in Table 1, the proposed datafication process consists three stages to quantify and convert various domains of urban social life, physical urban environment, and the perceptions of citizens to reach a data-driven understanding of citizens' social experience of urban environment.

The primary data collected from a behaviour mapping effort will allow us to capture the temporal and spatial distribution of the observed social interactions in urban public spaces. The social interactions will be mapped based on the different typologies, i.e., passive sociability, fleeting sociability, and enduring sociability (Mehta, 2019), within a variety of public spaces located in both the city centre and residential neighbourhood settings. In addition to digitising the data points, a spatial autocorrelation analysis will be conducted using ArcGIS to uncover the pattern by which citizens' social life unfolds in contemporary urban spaces.

A collection of open sourced data including smart cities open data (Sheffield City Council Open Data, 2023), Crowd sourced data, and authoritative online data will be batch processed using (open sourced GIS application) QGIS. We will conduct spatial statistical modelling to build an integrated urban model which evaluates the quality of the built environment. Specifically, the integrated model consists of three layers of geoprocessing functionality in terms of measuring the diversity (Shannon's Evenness Index), density (Ratio of employment density (population/workplace Points of Interests (POI)), and design (connectivity) of urban amenity /

infrastructure within the 800 m radius buffer of the observed public spaces (Kim & Hipp, 2021). The quality of the urban built environment in which the public spaces are situated in will be calculated by adopting a negative binomial regression approach (Kim & Hipp, 2021).

Finally, the online open-ended survey will collect primary data representing the demographics of members of the public and their views on: 1) physical characteristics of public spaces; 2) the kind of social interactions they experience; and 3) how they value their social experience in urban spaces. Using the statistical and text data collected from the online open-ended survey, we can validate the relationship between public spaces and social interaction by conducting moderation and mediation analysis using the sociodemographic and perceptual factors as the moderator and mediator. To further understanding citizens' perceptions of the social experience in urban environment, we will apply text mining and sentiment analysis to reveal their preferences of urban public spaces and values of social interactions using open sourced Python NLP libraries in terms of complementing the research findings.

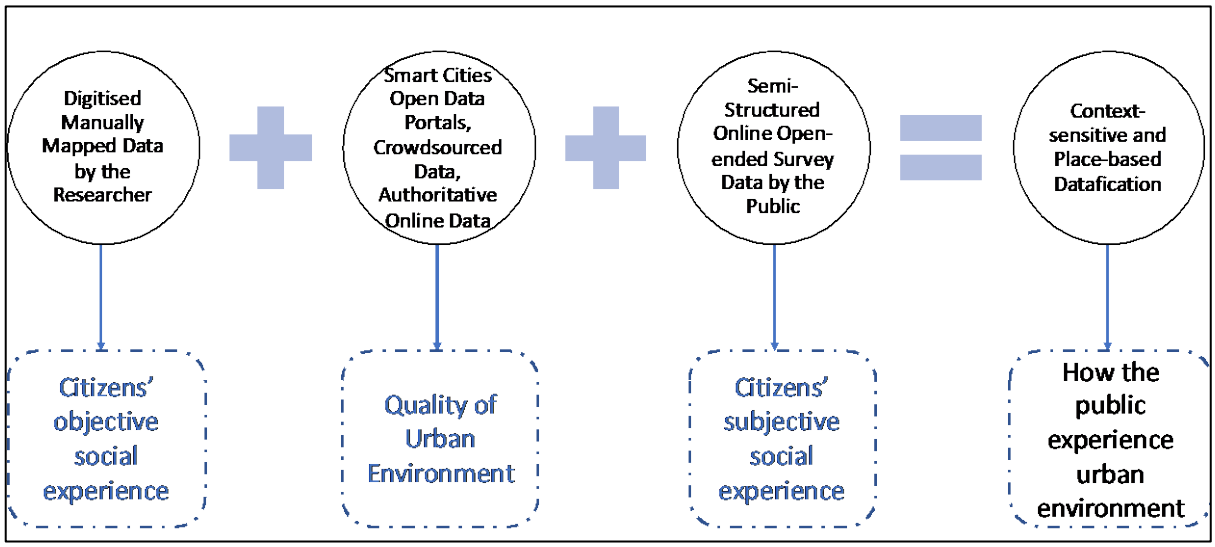

Figure 1: The complementary relationship of different data components.

As shown in Figure 1, the proposed research stages are complementary in terms of understanding how the public experiences urban environments by integrating data sourced from a behaviour mapping effort, online public participation, and across multiple open data platforms facilitated by the smart cities initiative. Such effort allows data-driven analytical thinking in smart sustainable urbanism practices (Bibri, 2019), as well as draws specially attention to the citizen-centric datafication that represents spatial realities of directly experienceable urban environment and contemporary social life (Christmann & Schinagl Leibniz, 2023). Specifically, the behaviour mapping process complies the physical and functional features of urban environments (Subiza-Pérez, Vozmediano, & Juan, 2019) whilst uncovering citizens' social experience within the urban setting using an objective approach. The open data we sourced from smart cities, authoritative, and crowd sourced platforms makes the basis of the raw data to develop urban space model that evaluates the quality of the built environment. In addition, the semi-structured data we collected via online open-ended survey not only enables smart citizen participation but also allows an in-depth understanding of the citizens' perceptions of their social experience in urban environment. The proposed research design, therefore, is closely aligned with the place-based and citizen-centric approach of exploring socially sustainable smart cities suggested by Aurigi & Odendaal (2021), by which we make priority-informed and data driven decisions utilising analytical techniques derived from urban informatics such as geoprocessing, statistical modelling, and text mining.

4. Conclusion

The construction of smart cities deemed to impact on the urban forms and everyday lifestyles with

placemaking becomes attentively integrated with pivotal technologies in equipping the datafication process in smart cities. In this study, we propose a datafication solution to understand citizens' lived experience in contemporary urban environments, proving the methodological insights which will help researchers and practitioners to better understand the social dimensions of smart cities development. As part of future work, we will look at understanding citizens' social life further in different city contexts (i.e., metro city, garden city,

megacity, etc) and further exploration in datafication solutions to transform domains of urban life and improve quality of urban environments,

References

Andreani, S., Kalchschmidt, M., Pinto, R., & Sayegh, A. (2019). Reframing technologically enhanced urban scenarios: A design research model towards human centered smart cities. *Technological Forecasting and Social Change*, 15-25.

Appio, F. P., Lima, M., & Paroutis, S. (2019). Understanding Smart Cities: Innovation ecosystems, technological advancements, and societal challenges. *Technological Forecasting & Social Change journal*, 1-14.

Aurigi, A., & Odendaal, N. (2021). From "Smart in the Box" to "Smart in the City": Rethinking the Socially Sustainable Smart City in Context. *Journal of Urban Technology*, 55-70.

Balsa-Barreiro, J., Menendez, M., & Morales, A. J. (2022). Scale, context, and heterogeneity: the complexity of the social space. *Nature Portfolio, Scientific Reports*, 1-12.

Bibri, S. E. (2019). The anatomy of the data driven smart sustainable city: instrumentation, datafication,

Holland, C., Clark, A., Katz, J., & Peace, S. (2007). Social interactions in urban public places. Joseph Rowntree Foundation by The Policy Press , 1-84.

Kim, Y. A., & Hipp, J. R. (2021). Density, diversity, and design: Three measures of the built environment and the spatial patterns of crime in street segments. *Journal of Criminal Justice*, 1-13.

Macke, J., Casagrande, R. M., Sarate, J. A., & Silva, K. A. (2018). Smart city and quality of life: Citizens' perception in a Brazilian case study. *Journal of Cleaner Production*, 717-726.

Mehta, V. (2019). Streets and social life in cities: a taxonomy of sociability. *Urban Design International*, 16-37.

Monfaredzadeh, T., & Krueger, R. (2015). Investing social factors in a smart city. *Procedia Engineering*, 1112-1118.

Nakano, S., & Washizu, A. (2021). Will smart cities enhance the social capital of residents? The importance of smart neighborhood management. *Cities*, 1-21.

Neves, F. T., Neto, M. C., & Aparicio, M. (2020). The impacts of open data initiatives on smart cities: A framework for evaluation and monitoring. Cities, 1-15.

Sheffield City Council Open Data. (2023, 7 5). Sheffield City Council Open Data. Retrieved from https://sheffield-city-council-open-data-sheffieldcc.hub.arcgis.com

Subiza-Pérez, M., Vozmediano, L., & Juan, C. J. (2019). A systematic social observation tool to measure the restorative potential of urban settings. *Psychology: Bilingual Journal of Environmental Psychology*, 257-267.

Trencher, G. (2019). Towards the smart city 2.0: Empirical evidence of using smartness as a tool for tackling social challenges. *Technological Forecasting & Social Change*, 117-128.

Zhao, Z., & Zhang, Y. (2020). Impact of Smart City Planning and Construction on Economic and Social Benefits Based on Big Data Analysis. *Complexity*, 1-11.

Late Submission