On small area composite indicators and classifications for urban planning: theory-driven and data-driven approaches

Alessia Calafiore
Edinburgh School of Architecture and Landscape Architecture, University of Edinburgh
acalafio@ed.ac.uk
orcid.org/0000-0002-5953-2891

Received: April 2025 / Accepted: December 2025 |
DOI: 10.36253/contest-15767

keywords
evidence based
urban planning
multidimensional indicators
GeoAI

The collection of data by public and private organisations has increased rapidly and continuously in the last decades. While some have called for "the end of theory" in a society dominated by data, this paper argues that theory and data driven approaches are both fundamental to make informed evidence-based decisions.
Taking the case of geographic data products aimed at supporting urban planning, this paper critically examines the approaches that underpin their development. Through illustrative examples, it describes two types of research data products – small-area

1. Introduction

Every second an immensely vast amount of data is collected via several means. This `big data` era has been driving significant changes in research and practice, opening up new ethical, methodological and analytical challenges.

Many of these big data are geographic, expressing the location of physical objects, humans and animals, often in real-time. The widespread presence of urban data sensors, sometimes directly linked to objects, as in the Internet of Things paradigm, and to people, as those voluntarily and involuntarily flowing through the use of mobile phones and their applications, have laid the foundations for the so-called smart cities. The idea of a smart city involves the creation of technological ecosystems whereby most of the urban everyday life processes can be governed by forms of automated decision-making (Ahad et al., 2020). While several cities around the

multidimensional theory-driven indicators and data-driven classifications – and outlines their respective advantages, limitations, and applications.
While it is recognised that geographic data products can be a valuable asset to support urban planning, challenges remain in translating research outputs into practice. To avoid technocratic and decontextualised applications of such data, it is suggested to prioritise reflexivity, situate knowledge, acknowledge uncertainty, and embrace openness throughout the data production and use process.

world have been implementing smart cities initiatives, the ambition to effectively manage urban complex dynamics by fully relying on technological supports have only partially been fulfilled and do not lack of criticism. For example, a relatively recent trend in the digital transformation towards smart cities is that of digital twins (Deng et al., 2021), which are defined as computer representations, or virtual copies, of the processes that determine how a physical system operates, connected by real-time data feeds with the physical system they represent (Batty, 2024). While digital twins have demonstrated potential in engineering applications and to describe physical phenomena, they are still not able to properly include people behaviours in their representations as well as the cultural, social and economic dynamics which are so essential to understand how cities function (Batty, 2024; Malleson et al., 2024).

More broadly, a criticism to the concept of smart cities, and related solutions such as the creation of digital twins, is that they are fundamentally grounded in a technocratic view to fixing urban problems (Malleson et al., 2024). Scholars who see urban planning theory and practice as the making of liveable, sustainable and healthy places, where people are at the very centre of the analysis, will certainly find it hard to believe that a purely technical solution can truly achieve the desired outcome. Urban planning that seeks to transform cities into better places to live must account for existing inequalities, risks of social exclusion, and people's lived experiences in order to achieve meaningful change.

In this contribution, I would argue that while existing urban models may not be good enough to find ready to implement solutions for place-making, the deluge of geographic data that we have seen flourishing in the last decades can certainly play a key role in transforming our cities by providing evidence urban planners can build upon.

To date, the collection, storing and analysis of geographic big data have greatly enriched our understanding of human and urban dynamics (Batty, 2013; Thakuriah et al., 2017), by generating new evidence on several phenomena such as spatial segregation dynamics, socio-spatial inequalities, human mobility patterns, the built environment impacts on health just to mention a few. Geographic big data have also been considered a fundamental asset in urban sustainability research, allowing to combine human behaviour, built environment and environmental data in the attempt to account for the complex and multidimensional nature of sustainable development at scale (Kong et al., 2020; Wang and Moriarty, 2018).

Leveraging new evidence to inform urban planning has been therefore increasingly a matter of discussion in the field, with scholars trying to position the role data and evidence may have to support the development of planning theories and practice. Krizek et al., (2009) highlighted the potential of evidence-based urban planning to bridge urban research and practice. At the same time, they called attention to the actual transferability of research results into practice, as well as to the importance of viewing evidence in a wider sense, therefore including participatory practice and accounting for the researchers positionality which may drive biases (Krizek et al., 2009).

Researchers creating Open Data Products (ODPs) take a step forward in favouring such transferability. ODPs can be thought as "the final data outcome resulting from adding value to raw, highly complex, unstructured and difficult-to-access data to address a well-defined problem, and making the generated data output openly available" (Arribas-Bel et al., 2021). Raw data are therefore analysed and reorganised to provide synthetic and more readable versions of the original sources and are developed to directly respond to certain research questions.

ODP when created with geographic data can respond to relevant questions about specific locations, for example how much walkable or bikeable an area is (Shashank and Schuurman, 2019, Lovelace et al., 2017), quantifying the level of access to services (Nicoletti et al., 2023, Calafiore et al., 2022) or sustainable urban mobility (Danielis et al. 2018), suggesting the extent to which the sustainable transport demand is met by the sustainable transport supply (Ballantyne et al., 2024), or what the main features of an area are, either in terms of their morphology (Alexiou et al., 2016, Fleishmann et al. 2025) and/or the land use (Arribas-Bel and Fleischmann, 2022; Samardzhiev et al., 2022).

Although the creation of geographic data products is increasingly common in urban planning research, how they can become an integrated part of urban planning practice is still an open question. This paper aims at first, describing the advantages and limitations of geographic data products by critically discussing two epistemologically distinct perspectives in their making - theory-driven and data-driven; second, highlighting current challenges on how these research data can be adopted in planning practice.

2. Creating human readable data

Creating multidimensional indicators and classifications can be described as a series of analytical tasks that transform raw data into data products, where raw data are not interpretable by humans while data products are. Therefore, they have the potential to be a valuable resource to inform decision-making. In this contribution two widely used and distinct approaches to build data products at small area level are discussed.

2.1 Small area composite indicators

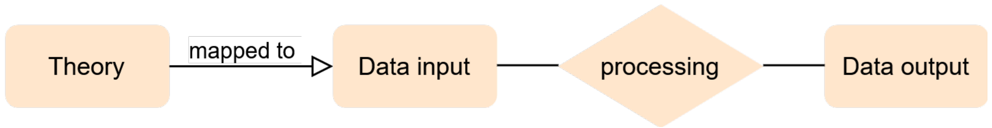
Composite indicators emerge from the need to describe a complex multidimensional phenomenon with a synthetic measure, rather than forcing a decision-maker to look at multiple indicators individually (Mazziotta and Pareto, 2017). As the volume of available data grows, so does the opportunity to combine information from multiple sources, making the development of analytical tools – such as composite indicators or indices – increasingly common.

Several international organisations make use of composite indicators and contributed to standardise and support its development with handbooks, i.e. the Organisation for Economic Co-operation and Development handbook to construct composite indicators (Nardo et al., 2008), or through working groups, i.e. the United Nation Statistics Division Working Group on Composite Indices[1].

While many existing composite indicators are related to entire countries, urban planning is concerned with how cities and places function, therefore the granularity of the data's spatial dimension becomes key to inform it. Consequently, this contribution focuses on those which provide synthetic information of small geographical areas, although many of the discussion points are generalisable to all composite indicators.

At their core, these indicators are a statistical aggregation of data that capture the various dimensions of a phenomenon in a synthetic way. A fundamental step in constructing composite indicators is the development of a theoretical framework to define the multidimensional phenomenon of interest. It is the central importance of this step that distinguishes composite indicators from multidimensional classifications. The methodological choices that follow are, along with statistical validity, intended to best represent the underlying theoretical framework.

When defining a theoretical framework, the data creator must make three key decisions: identifying the domain and its sub-domains, specifying their properties, and determining



the relative importance of those properties. The domain represents the highest level of analysis, capturing the core characteristics of the phenomenon of interest, while sub-domains allow this phenomenon to be broken down into meaningful components. Properties describe the essential features needed to characterise the phenomenon. Their importance may be equal or vary across domains and sub-domains, and this variation is typically expressed through the use of weights or hierarchical structures.

It is important to note that composite indicators are ultimately made of numbers, therefore limited by the measurability of a property or by the lack of data. Ideally, the initial definition of the theoretical framework should go beyond these limitations (Nardo et al., 2008) in order to make explicit what is not captured. However, a clear mapping between the identified properties and the data is also fundamental for the interpretability of composite indicators (see Figure 1).

Theoretical frameworks are grounded in existing theories, ideally backed by evidence, that characterise the domain of interest. In many cases, the resulting theoretical framework is far from being universally recognised with scholars often debating the very definition of a phenomenon and evaluating the addition or deletion of certain properties as well as the varying or their relative importance. An example of composite indicator in the context of urban planning that has been highly debated in the last decades is the walkability index. Extensive literature review shows how the concept of walkability and the way it has been measured is still part of a broad discussion around what it means for a place to be walkable (Shields et al., 2023). However, the most established and common theoretical frameworks to develop walkability indices are influenced by the 3 Ds (Diversity, Density and Design) theory, then extended to the 5 Ds (Destination accessibility and Distance to transit) theory (Venerandi et al., 2024). Brought about by new urbanists, the 3Ds theory suggests that a compact built environment with a mixed land use and a good design can reduce motorised travel, which has also been confirmed by evidence leveraging data (Cervero and Kockelman, 1997). This principle has been translated into walkability indices by using variables that convey the 3 Ds as in Frank et al. (2005) combining residential density, land use mix and street connectivity. Building on Frank et al. (2005), variables mirroring Destination accessibility and Distance to transit have been included by some other walkability indices as in Neckerman et al. (2009) where distance to subways and retail areas are accounted for.

Theoretical frameworks can also be informed by views emerging from the engagement of the public. This approach is more common in applied research or practice, where a non-

High level representation of composite indicators development.
Source: made by the author
Fig. 1

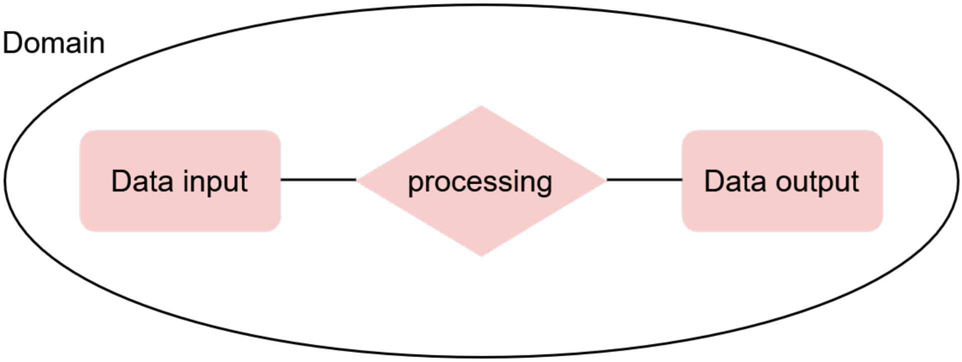
academic perspective plays an important role for the validity and acceptability of the theoretical framework, especially when the composite indicator is directly meant to support policy making. The importance of engaging the public lies primarily in the need to incorporate bottom-up information that more accurately reflects people's views and lived experiences. This is evident in walkability indices that integrate perceived walkability through surveys (Zhang and Mu, 2020), rely on interviews or fieldwork (Knapskog et al., 2019), or in accessibility indices that consider the preferences of specific population groups—for example, older adults (Dunning et al., 2023). Along with bottom-up knowledge, expert knowledge coming from non-academic stakeholders can also contribute to developing a theoretical framework more targeted on making impact on and influence decision-making (Ballantyne and Singleton, 2024).

2.2 Small area classifications

While small area classifications for urban planning are not built by ignoring urban theories and dynamics, compared to combined indicators they are mostly aimed at knowledge discovery `from` the data rather than capturing existing knowledge `through` the data. Creating small area classification relies on machine learning methods that identify clustering patterns across multidimensional data. The growing intersection between Data Science and Geographic Information Science — driven by the opportunities offered by geographic big data — has been widely discussed over the past decade, with many calling for deeper integration between the two fields (Singleton and Arribas-Bel, 2021). Small-area classifications provide a compelling example of this ongoing cross-pollination.

Clustering algorithms developed in data science as unsupervised machine learning approaches are now used in many geographic applications. Broadly, a clustering task aims at dividing data points into groups such that the similarity among those in the same group and the dissimilarity between data points in different groups are maximised. When applied to spatial data this task generates new geographies of the urban environment, where space is divided based on patterns emerging from data. Such exercise has been widely implemented to discover population geographies also known as geodemographics by characterising people living in the same area; to synthetically describe the physical urban environment by characterising the built environment morphology in the same area; or to classify places more broadly by accounting for both physical and population characteristics in the same area.

While combined indicators are designed to represent a well framed urban phenomenon, small area classifications tend to be more general purpose and then applied in different ways.



The development of geodemographics have found several applications to support urban planning. Broadly speaking, geodemographic classifications offer a condensed, multidimensional description of the population living in an area and can be used to better understand existing inequalities among segments of the population rather than focusing on individual variables. For example, they can undercover inequalities on who is getting access to green spaces (Barbosa et al., 2007), to the urban amenities people need (Dunning et al., 2023), or even to assess how successful the spatial targeting of urban policy initiatives has been for the benefit of certain segments of the population (Batey and Brown, 2007).

Along with mapping population groups, small area classifications can provide an informative layer describing the forms and functions of the built environment. Land cover and land use maps have been part of the urban planner toolkit for decades; with the increasing amount of spatial data at a more and more granular resolution, a new generation of small area classifications of the urban fabric is emerging. These kinds of classifications use a wider variety of data input that enable more flexibility while providing descriptions of the built environment, by developing novel spatial units to better capture variations in urban forms and functions (Arribas-Bel and Fleischmann, 2022), by focus on specific types of infrastructure such as green infrastructures (Morpurgo et al., 2023), or analyse them in combination with geodemographics (Alexiou et al., 2016).

As mentioned in the beginning of this section, small area classifications are mostly aimed at knowledge discovery driven by data patterns learned by algorithms; however, there is a theoretical component behind the development of these classifications in the data selection stage. Selecting which data to include or to exclude requires an understanding of the phenomenon that you want to discover through data. On the one hand, since machine learning applications do benefit from the quantity, along with the quality, of data, the data selection tends to be less restrictive than in composite indicators, and it is mostly focused at capturing the general domain of interest to then discover knowledge rather than to align the data to existing theory and knowledge as for composite indicators. On the other hand, this theoretical component opens up the opportunity to also engage the public in the data selection process; an interesting

High level representation of multidimensional classifications development
Source: made by the author
Fig. 2

example is the ageing in place classification developed by (Yang et al., 2023) for which consultations with stakeholders were made in the data selection phase to obtain feedback on the relevance of the included variables. In Fleischmann et al. (2025), although there is no direct mapping of the theoretical framework to the data input, the role of theory goes beyond data selection, to inform methods that defines urban forms.

3. Comparing small-area indicators and classifications

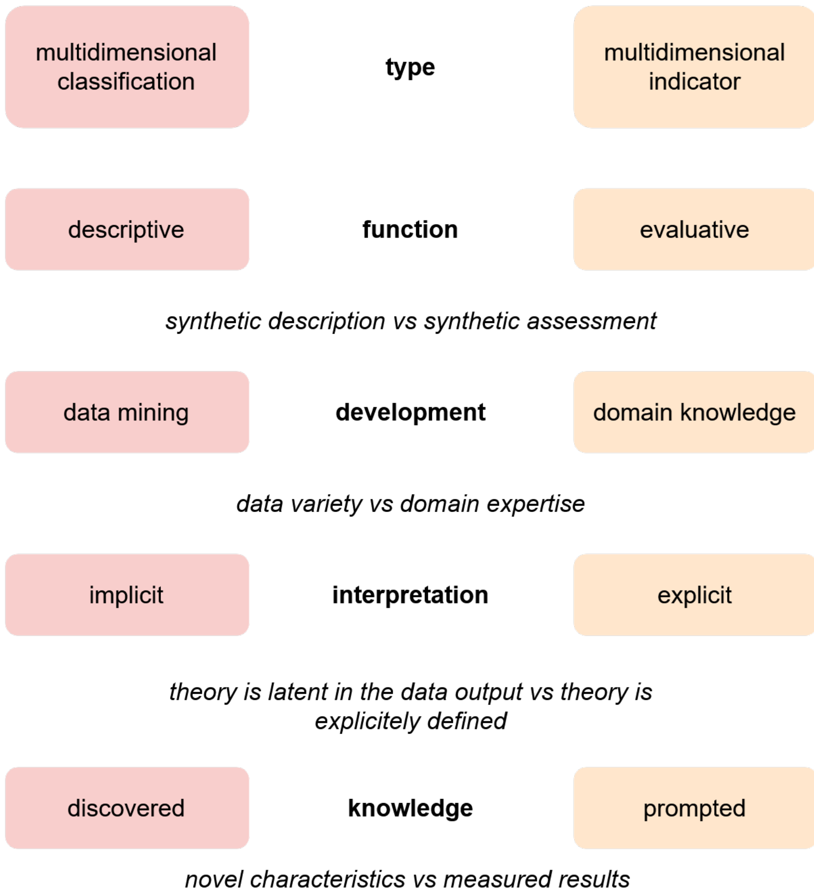
Cities are complex systems arising from the mutual interaction between humans and the built environment. While data offer new opportunities to better understand and eventually govern urban dynamics to make cities more liveable and sustainable, cross-fertilisation between data science expertise and urbanism is needed to unlock the urban data potential in supporting decision-making.

In this contribution, I have presented two types of geographic data products that offer strong potential for making emerging evidence more intelligible to users while also aligning more closely with urban planning theory and practice.

Both composite indicators and classifications are multidimensional, capturing a phenomenon from different perspectives with only one measure. This is an advantage from at least two viewpoints. First, studies have shown that composite metrics tend to be more robust after validation than single indicators (Shashank and Schuurman, 2019); second, they provide a more comprehensive description of a place that is more easily legible than variables taken individually. Concerning this latter aspect, it is important to note a difference between composite indicators and classifications: the former can be seen as evaluative, it aims at answering questions related to the `how much` of a specific phenomenon, i.e. how much sustainable, liveable, walkable, vulnerable an area is; the latter have a more descriptive purpose and require additional explanations that render them more legible to users such as `pen portraits` comprising text and illustrative material summarising the prevailing characteristics of groups and types (Longley and Singleton, 2009).

As shown in Figures 1 and 2 multidimensional indicators and classifications follow distinct development workflows. The development of composite indicators corresponds to a theory-driven approach, relying largely on the creator's domain expertise to define how areas are characterised and, importantly, ranked or evaluated. In contrast, the development of classifications – while still informed by theory – focuses more heavily on drawing from a broader range of data to uncover meaningful patterns through data-mining techniques that reveal hidden structures.

From a theoretical perspective the



interpretation of composite indicators and multidimensional classifications come at different moments: composite indicators should be grounded in explicit theoretical frameworks that can be easily situated and critically assessed by the users even before looking at the resulting maps; multidimensional classifications cannot be scrutinised a-priori, making the interpretation slightly more complex, as it comes afterwards while assessing the results. Such difference has a methodological counterpart which results in distinct kinds of outcome. Composite indicators require methods that aim at combining data as much representative as possible of the theoretical framework, obtaining information about an area that is prompted by such framework; multidimensional classifications result from methods that are thought to be the best fit for mining the available data, generally unsupervised, and obtaining new, data-driven, description of an area.

While neither composite indicators or

Comparing small-area multidimensional classifications and indicators.
Source: made by the author
Fig. 3

multidimensional classifications are intrinsically better than the others, it is crucial to appreciate their differences and make a context dependent use of them. Cases where a scholar or practitioner has a question on a very specific urban phenomenon and can easily access domain expertise on such phenomenon should make use of composite indicators; on the contrary, if the concern is on exploring the characteristics of places more broadly, multidimensional classifications would be much more useful and allow to make knowledge emerge from data.

4. From research to practice: challenges and opportunities

Composite indicators and multidimensional classifications at small area level are widely employed in research; however, they are not often applied to real-word urban planning practice. While data-driven solutions in the context of smart cities often materialise as part of a technocratic and top down view of city governance, the advantage of legible and easy to visualise geographic data products is that they can become a valuable tool for practitioners and policy makers to not only develop evidence based urban plans, but also to communicate and interact with the local community about such plans in a more informative and accountable way. As van Dijk et al. (2020) put it the ambition for evidence-based urban planning is that of creating the basis for "an informed and empowered community to engage democratically in the local administration of our cities and neighbourhoods".

However, to maximise the opportunities of transferring research outputs into real world urban planning, it is argued that there are still some challenges that must be addressed on the way such outputs are developed and packaged. Notably, it is suggested to prioritise reflexivity, situate knowledge, acknowledge uncertainty, and embrace openness throughout the data production and use process.

4.1 Reflexivity

Data are never neutral and can be thought as "socially constructed artefacts that reflect the contexts and processes of their creation" (Iliadis and Russo, 2016). Composite indicators result from the way a phenomenon of interest is defined by the data product creator and their underlying assumptions (Shashank and Schuurman, 2019); likewise, interpreting results to provide legible descriptions from classifications cannot be seen as objective but influences by the data creator positionality. Both these aspects need to be taken into consideration while interacting with any type of data products. It therefore becomes fundamental to accompany geographic data products with digestible pieces of information making theoretical choices, assumptions and possible source of biases explicit.

4.2 Situating knowledge

Urban contexts vary substantially, and their descriptions require an in-depth understanding of their history, culture and socio-economic structure. The increasing availability of data globally renders the development of indicators and classifications for cities across the world now possible. While this is certainly an advantage to generate far-reaching geographic data products and enable comparisons between countries, it risks excluding locally relevant perspectives in their making. Because power dynamics shape both people and data, a Western-centric perspective on urban planning can easily permeate the development of geographic data products, excluding relevant views in the making. Moreover, research shows that biases rooted in structural inequalities are often reproduced in the data itself, leading to the over-representation of certain areas or population groups at the expense of others (Robinson and Franklin, 2021). One avenue that could be pursued moving forward would involve the data creator providing a degree of flexibility in the outputs, for instance by decomposing the data product into constituent components that can be recombined and augmented with locally specific knowledge.

4.3 Uncertainty

As Couclelis (2003) puts it, GIS and geographic knowledge are limited by the certainty of uncertainty. Some of the errors and uncertainties can be attributable to the specificities of geospatial data processing such as the use of different spatiotemporal scales and zonal schemes or the definition of geographic context in space and time (Chun et al., 2019). The Modifiable Area Unit Problem (MAUP) is a well-known problem in geographic data processing whereby the definition of the area unit determines the way a phenomenon is represented, and its variation can bring inconsistent results (Wong, 2004). Kwan (2012) highlights that, especially when dealing with place-people relationships, more dynamic area unit definitions across space and time are needed to overcome what she called the Uncertain Geographic Context Problem (UGCoP). While some techniques can minimise the propagation of errors and uncertainty, these remain an intrinsic characteristic of geographic knowledge (Couclelis, 2003). To make geographic data products that can be reliably used in urban planning practice is therefore crucial to be explicit on the underlying uncertainty, how this has been mitigated and what are the limitations in the data employed. Furthermore, I would argue that significant work remains in bridging qualitative and quantitative approaches to address the knowledge gaps arising from the loss of non-measurable information when relying exclusively on data, thereby enabling more informed decision-making.

4.4 Openness and transparency

Openness and transparency are fundamental to enable the reuse of geographic data products both in research and practice. While not all input data can always be publicly shared, pushing towards opening the derived data products as pure datasets or through interactive visualisations is certainly an effective way to enable their use in real world urban planning scenarios. However, making data products publicly accessible does not come without challenges. The FAIR (Findable, Accessible, Interoperable, Reusable) principles provide guidelines to researchers on how to publish their research data and serve as a valuable example of how advocating for data sharing is becoming increasingly common (Wilkinson et al., 2016); academic journals are also growingly recognising the value of sharing research data and creating outlets, i.e. Nature Scientific Data, or paper types, i.e. urban data/code (Arribas-Bel et al., 2021), specifically meant at research data publishing. Although making research data more accessible is taking on an increasingly central role, how to make them accessible beyond academia is still not broadly discussed. Notably, accompanying geographic data product with dissemination-oriented documentation and designing interactive visualisation could be among best practices moving towards a more outreach focused data publishing.

5. Conclusion

In the last decades, data have become a fundamental resource in all sectors of our society, including the planning of cities. However, this brings about several challenges encompassing the theoretical, methodological and ethical sphere.

In this contribution, two types of rather common data products that can support decision-making in urban planning have been presented and compared. Notably, these two have been selected to contrast the theory-driven approach of composite indicators and the data-driven approach of multidimensional classifications. The two approaches exemplify some of the challenges that we broadly see while using data with new AI methods in the social domain.

The functioning of our societies and therefore of our cities have been object of studies way before the technologies we make use of today were even part of our imagination. The role of urban theory and the importance to build on existing knowledge coming from centuries of observations of cities and their dynamics should never be underestimated. Interdisciplinary work and cross-fertilisation between urbanism and data science need to become more common to avoid the development of a technocratic vision of cities, where data products are not properly interpreted and challenged.

At the same time, geographic big data allow to scale observations up to an unprecedented level and this is an unmissable opportunity to discover new knowledge about urban dynamics. AI methods often develop solutions coming from a so-called black-box where the rationale behind a machine decision is not known. Since a city is not an engine and there is no unique optimal solution to make it a better place to live in, urban planners have to motivate their decisions which may not be the best solution in absolute terms, but what a planner believes being the best in a certain context or based on a certain vision; as a consequence, to inform decision-making with data-driven evidence there need to be a comprehensive understanding of what the determinants of such evidence are, which is still not always possible with AI. A promising development of AI in this direction is explainable AI that aims at providing metrics to describe how a machine solved a e.g. classification problem, including which one of the data input have played a more significant role. Furthermore, in light of bridging urban planning with AI there need to be an effort in providing not only more accessible data products but also more accessible documentations on data quality and awareness of the intrinsic uncertainty of any quantitative science dealing with human dynamics[2].

It is also important to note that some of the uncertainty we have to deal with using data is also due to privacy concerns. Ethical data collection and usage are still a very open issue since AI systems came into play. The combination of increasingly personal information people give freely to big platforms with the growing computation capability generates significant ethical concerns. Such concerns invest data that can be used to inform urban planning too. There are already a multitude of studies that made use of the so-called 'new forms of data', which include those coming from relatively new sources such as social media or mobile phone applications. These new forms of data inevitably open up new opportunities to fill that uncertainty gap with more and more detailed information on how people navigate city spaces and their preferences (Ki et al., 2023). While this is certainly an advantage for urban research, ensuring that there is ongoing and explicit consents from data subjects is still an open issue.

Notes

[1] https://unstats.un.org/unsd/methods/wgci/default.htm

[2] Franklin (2022) defines uncertainty "broadly as the gap that exists between real-world, 'true but unknown' values and relationships and what we are able to observe, given the methods and data available."

References

Ahad, M.A., Paiva, S., Tripathi, G., Feroz, N., 2020. Enabling technologies and sustainable smart cities. Sustain. Cities Soc. 61, 102301. https://doi.org/10.1016/j.scs.2020.102301

Alexiou, A., Singleton, A., Longley, P.A., 2016. A Classification of Multidimensional Open Data for Urban Morphology. Built Environ. 42, 382–395. https://doi.org/10.2148/benv.42.3.382

Arribas-Bel, D., Fleischmann, M., 2022. Spatial Signatures - Understanding (urban) spaces through form and function. Habitat Int. 128, 102641. https://doi.org/10.1016/j.habitatint.2022.102641

Arribas-Bel, D., Green, M., Rowe, F., Singleton, A., 2021. Open data products-A framework for creating valuable analysis ready data. J. Geogr. Syst. 23, 497–514. https://doi.org/10.1007/s10109-021-00363-5

Arribas-Bel, D., Alvanides, S., Batty, M., Crooks, A., See, L., & Wolf, L. 2021. Urban data/code: a new EP-B section. Environment and Planning B: Urban Analytics and City Science, 48(9), 2517-2519.

Ballantyne, P., Filomena, G., Rowe, F., Singleton, A., 2024. Developing two-dimensional indicators of transport demand and supply to promote sustainable transportation equity. Comput. Environ. Urban Syst. 113, 102179. https://doi.org/10.1016/j.compenvurbsys.2024.102179

Ballantyne, P., Singleton, A., 2024. Using composite indicators and city dashboards to promote place-based policy interventions. Cities 154, 105329. https://doi.org/10.1016/j.cities.2024.105329

Barbosa, O., Tratalos, J.A., Armsworth, P.R., Davies, R.G., Fuller, R.A., Johnson, P., Gaston, K.J., 2007. Who benefits from access to green space? A case study from Sheffield, UK. Landsc. Urban Plan. 83, 187–195. https://doi.org/10.1016/j.landurbplan.2007.04.004

Batey, P., Brown, P., 2007. The Spatial Targeting of Urban Policy Initiatives: A Geodemographic Assessment Tool. Environ. Plan. Econ. Space 39, 2774–2793. https://doi.org/10.1068/a38519

Batty, M., 2024. Digital twins in city planning. Nat. Comput. Sci. 4, 192–199. https://doi.org/10.1038/s43588-024-00606-7

Batty, M., 2013. Big data, smart cities and city planning. Dialogues Hum. Geogr. 3, 274–279. https://doi.org/10.1177/2043820613513390

Calafiore, A., Dunning, R., Nurse, A., Singleton, A., 2022. The 20-minute city: An equity analysis of Liverpool City Region. Transp. Res. Part Transp. Environ. 102, 103111. https://doi.org/10.1016/j.trd.2021.103111

Calafiore, A., Palmer, G., Comber, S., Arribas-Bel, D., Singleton, A., 2021. A geographic data science framework for the functional and contextual analysis of human dynamics within global cities. Comput. Environ. Urban Syst. 85, 101539. https://doi.org/10.1016/j.compenvurbsys.2020.101539

Chun, Y., Kwan, M. P., & Griffith, D. A., 2019. Uncertainty and context in GIScience and geography: challenges in the era of geospatial big data. International Journal of Geographical Information Science, 33(6), 1131-1134.

Couclelis, H. 2003. The certainty of uncertainty: GIS and the limits of geographic knowledge. Transactions in GIS, 7(2), 165-175.

Cervero, R., Kockelman, K., 1997. Travel demand and the 3Ds: Density, diversity, and design. Transp. Res. Part Transp. Environ. 2, 199–219. https://doi.org/10.1016/S1361-9209(97)00009-6

Danielis, R., Rotaris, L. and Monte, A., 2018. Composite indicators of sustainable urban mobility: Estimating the rankings frequency distribution combining multiple methodologies. International Journal of Sustainable Transportation, 12(5), pp.380-395.

Deng, T., Zhang, K., Shen, Z.-J. (Max), 2021. A systematic review of a digital twin city: A new pattern of urban governance toward smart cities. J. Manag. Sci. Eng. 6, 125–134. https://doi.org/10.1016/j.jmse.2021.03.003

Dunning, R.J., Dolega, L., Nasuto, A., Nurse, A., Calafiore, A., 2023. Age and the 20-min city: Accounting for variation in mobility. Appl. Geogr. 156, 103005. https://doi.org/10.1016/j.apgeog.2023.103005

Engin, Z., van Dijk, J., Lan, T., Longley, P. A., Treleaven, P., Batty, M., & Penn, A. 2020. Data-driven urban management: Mapping the landscape. Journal of Urban Management, 9(2), 140-150. https://doi.org/10.1016/j.jum.2019.12.001

Fleischmann, M., Samardzhiev, K., Brázdová, A., Dančejová, D. and Winkler, L., 2025. The Hierarchical Morphotope Classification: A Theory-Driven Framework for Large-Scale Analysis of Built Form. arXiv preprint arXiv:2509.10083.

Frank, L.D., Schmid, T.L., Sallis, J.F., Chapman, J., Saelens, B.E., 2005. Linking objectively measured physical activity with objectively measured urban form. Am. J. Prev. Med. 28, 117–125. https://doi.org/10.1016/j.amepre.2004.11.001

Franklin, R., 2022. Quantitative methods I: Reckoning with uncertainty. Prog. Hum. Geogr. 46, 689–697. https://doi.org/10.1177/03091325211063635

Iliadis, A., Russo, F., 2016. Critical data studies: An introduction. Big Data Soc. 3, 2053951716674238. https://doi.org/10.1177/2053951716674238

Ki, D., Chen, Z., Lee, S., Lieu, S., 2023. A novel walkability index using google street view and deep learning. Sustain. Cities Soc. 99, 104896. https://doi.org/10.1016/j.scs.2023.104896

Knapskog, M., Hagen, O.H., Tennøy, A., Rynning, M.K., 2019. Exploring ways of measuring walkability. Transp. Res. Procedia 41, 264–282. https://doi.org/10.1016/j.trpro.2019.09.047

Kong, L., Liu, Z., Wu, J., 2020. A systematic review of big data-based urban sustainability research: State-of-the-science and future directions. J. Clean. Prod. 273, 123142. https://doi.org/10.1016/j.jclepro.2020.123142

Krizek, K., Forysth, A., Slotterback, C.S., 2009. Is There a Role for Evidence-Based Practice in Urban Planning and Policy? Plan. Theory Pract. 10, 459–478. https://doi.org/10.1080/14649350903417241

Kwan, M. P., 2012. The uncertain geographic context problem. Annals of the Association of American Geographers, 102(5), 958-968.

Longley, P.A., Singleton, A.D., 2009. Classification through consultation: public views of the geography of the e-Society. Int. J. Geogr. Inf. Sci. 23, 737–763. https://doi.org/10.1080/13658810701704652

Lovelace, R., Goodman, A., Aldred, R., Berkoff, N., Abbas, A. and Woodcock, J., 2017. The Propensity to Cycle Tool: An open source online system for sustainable transport planning. Journal of transport and land use, 10(1), pp.505-528. https://www.jstor.org/stable/26211742

Malleson, N., Franklin, R., Arribas-Bel, D., Cheng, T., Birkin, M., 2024. Digital twins on trial: Can they actually solve wicked societal problems and change the world for better? Environ. Plan. B Urban Anal. City Sci. 23998083241262893. https://doi.org/10.1177/23998083241262893

Mazziotta, M., Pareto, A., 2017. Synthesis of Indicators: The Composite Indicators Approach, in: Complexity in Society: From Indicators Construction to Their Synthesis, Social Indicators Research Series. Springer International Publishing, Cham, pp. 159–192. https://doi.org/10.1007/978-3-319-60595-1

Morpurgo, J., Remme, R.P., Van Bodegom, P.M., 2023. CUGIC: The Consolidated Urban Green Infrastructure Classification for assessing ecosystem services and biodiversity. Landsc. Urban Plan. 234, 104726. https://doi.org/10.1016/j.landurbplan.2023.104726

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., Giovannini, E., 2008. Handbook on constructing composite indicators: Methodology and user guide. OECD publishing.

Nicoletti, L., Sirenko, M., & Verma, T., 2023. Disadvantaged communities have lower access to urban infrastructure. Environment and Planning B: Urban Analytics and City Science, 50(3), 831-849. https://doi.org/10.1177/23998083221131044

Neckerman, K.M., Lovasi, G.S., Davies, S., Purciel, M., Quinn, J., Feder, E., Raghunath, N., Wasserman, B., Rundle, A., 2009. Disparities in Urban Neighborhood Conditions: Evidence from GIS Measures and Field Observation in New York City. J. Public Health Policy 30, S264–S285. https://doi.org/10.1057/jphp.2008.47

Robinson, C., Franklin, R.S., 2021. The sensor desert quandary: What does it mean (not) to count in the smart city? Trans. Inst. Br. Geogr. 46, 238–254. https://doi.org/10.1111/tran.12415

Samardzhiev, K., Fleischmann, M., Arribas-Bel, D., Calafiore, A., Rowe, F., 2022. Functional signatures in Great Britain: A dataset. Data Brief 43, 108335. https://doi.org/10.1016/j.dib.2022.108335

Shashank, A., Schuurman, N., 2019. Unpacking walkability indices and their inherent assumptions. Health Place 55, 145–154. https://doi.org/10.1016/j.healthplace.2018.12.005

Shields, R., Gomes Da Silva, E.J., Lima E Lima, T., Osorio, N., 2023. Walkability: a review of trends. J. Urban. Int. Res. Placemaking Urban Sustain. 16, 19–41. https://doi.org/10.1080/17549175.2021.1936601

Singleton, A., Arribas-Bel, D., 2021. Geographic Data Science. Geogr. Anal. 53, 61–75. https://doi.org/10.1111/gean.12194

Thakuriah, P., Tilahun, N.Y., Zellner, M., 2017. Big Data and Urban Informatics: Innovations and Challenges to Urban Planning and Knowledge Discovery, in: Thakuriah, P., Tilahun, N., Zellner, M. (Eds.), Seeing Cities Through Big Data, Springer Geography. Springer International Publishing, Cham, pp. 11–45. https://doi.org/10.1007/978-3-319-40902-3_2

Venerandi, A., Mellen, H., Romice, O., Porta, S., 2024. Walkability Indices—The State of the Art and Future Directions: A Systematic Review. Sustainability 16, 6730. https://doi.org/10.3390/su16166730

Wang, S.J., Moriarty, P., 2018. Big Data for Urban Sustainability. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-73610-5

Wong, D. W., 2004. The modifiable areal unit problem (MAUP). In WorldMinds: geographical perspectives on 100 problems: commemorating the 100th anniversary of the association of American geographers 1904–2004 (pp. 571-575). Dordrecht: Springer Netherlands.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. 2016. The FAIR Guiding Principles for scientific data management and stewardship. Scientific data, 3(1), 1-9. https://doi.org/10.1038/sdata.2016.18

Yang, Y., Dolega, L., Darlington-Pollock, F., 2023. Ageing in Place Classification: Creating a geodemographic classification for the ageing population in England. Appl. Spat. Anal. Policy 16, 583–623. https://doi.org/10.1007/s12061-022-09490-y

Zhang, X., Mu, L., 2020. The perceived importance and objective measurement of walkability in the built environment rating. Environ. Plan. B Urban Anal. City Sci. 47, 1655–1671. https://doi.org/10.1177/2399808319832305