Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Co-creating data ecosystems with urban informal communities for justice and transformation

Joao Porto de Albuquerque

Joao.Porto@glasgow.ac.uk

University of Glasgow https://orcid.org/0000-0002-3160-3168

Angela Abascal

Public University of Navarra

Olubunmi Alugbin

University of Lagos

Taiwo Amole

Bayero University Kano

Kehinde Baruwa

University of Lagos

Caleb Cheruiyot

African Population and Health Research Centre

Andrew Mitham-Clarke

University of Glasgow

Darren Dixon

University of Glasgow

Xingyi Du

University of Glasgow

Peter Elias

University of Lagos

Helen Elsey

University of York

Ryan Engstrom

George Washington University

Grace Gielink

University of Glasgow

Serkan Girgin

Faculty of Geo-Information Science and Earth Observation, University of Twente
https://orcid.org/0000-0002-0156-185X

Sebastian Hafner

University of Glasgow

Esther Judith
African Population and Health Research Centre
Diego Pajarito-Grajales
University of Glasgow
Caroline Kabaria
African Population and Health Research Centre
Monika Kuffer
twente https://orcid.org/0000-0002-1915-2069
Robert Ndugwa
UN-Habitat https://orcid.org/0000-0003-3286-9061
Oluwatoyin Odulana
University of Lagos
Francis Onyambu
African Population and Health Research Centre
Sabitu Sa'adu Da'u
Africa Center of Excellence for Population Health and Policy and Department of Community Medicine, Bayero University Kano
Oluwatimilehin Shonowo
University of Glasgow
Dana R. Thomson
Columbia University
Grant Tregonning
University of Glasgow
Qunshan Zhao
University of Glasgow https://orcid.org/0000-0002-5549-9457
Mingyu Zhu
University of Glasgow

Article

Keywords:

Posted Date: March 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8927918/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Co-creating data ecosystems with urban informal communities for justice and transformation

João Porto de Albuquerque*, Angela Abascal, Olubunmi Alugbin, Taiwo Gboluwaga Amole, Kehinde Baruwa, Caleb Cheruiyot, Andrew Mitham-Clarke, Darren Dixon, Xingyi Du, Peter Elias, Helen Elsey, Ryan Engstrom, Grace Gielink, Serkan Girgin, Sebastian Hafner, Esther Judith, Diego Pajarito Grajales, Caroline Kabaria, Monika Kuffer, Robert Ndugwa, Oluwatoyin Odulana, Francis Onyambu, Sabitu Sa'adu Da'u, Oluwatimilehin Adenike Shonowo, Dana R. Thomson, Grant Tregonning, Qunshan Zhao, Mingyu Zhu.

Abstract
Artificial Intelligence and Earth Observation offer powerful tools for advancing urban science; however, they risk perpetuating social and epistemic injustices when data misrepresent deprived urban areas, such as slums and informal settlements. In response, this study introduces the IDEAMAPS Data Ecosystem, a transdisciplinary framework for co-creating justice-oriented urban data. An innovative digital collaboration platform and an interactive process support the co-creation of AI-based urban models grounded in communities' priorities and assets. In partnership with local authorities, data scientists, and global experts, our methodology enables a more precise and action-oriented multidimensional characterisation of deprived areas. Evaluations across the study sites in Nairobi, Lagos, and Kano demonstrate that community-grounded data ecosystems transform both the accuracy and the politics of modelling urban informality. These findings show how community involvement can systematically enhance urban models, thereby bolstering participatory urban regeneration initiatives.

1. Introduction

Recent advances in Artificial Intelligence (AI) have sparked both enthusiasm over their transformative potential[1,2] and concerns about the risks of bias, harm, and adverse consequences[3]. While AI encompasses a broad range of methods and applications, recent discussions have largely centred on large-language models such as OpenAI's ChatGPT and Google Gemini. Beyond these models, integrating AI with geospatial and Earth Observation (EO) data, such as satellite and street-level imagery, offers unprecedented opportunities to derive new insights for understanding and improving urban health, environmental conditions, and wellbeing[4,5]

This potential is especially significant in the Global South, where over 1.1 billion people live in informal settlements and underserved urban areas known as slums[6]. These areas are often underrepresented in data systems, creating a major obstacle to understanding the social and environmental factors that influence health and urban inequality[7]. AI-powered urban science models can produce fine-grained data on deprived areas, helping close these critical data gaps that hinder progress toward global policy targets, such as the Sustainable Development Goals (SDGs)[8]. As highlighted in the 2030 Agenda for Sustainable Development[9], "quality, accessible, timely and reliable disaggregated data" are essential to ensuring that no one is left behind, though such data remain largely scarce. Access to fine-grained, up-to-date

data is key to decision-making, enabling more effective, targeted interventions and helping identify and address inequalities, social exclusion, and discrimination.

However, the inequalities intended to be addressed also pose two crucial challenges and risks for the use of AI and data analytics for sustainable urban development. First, social inequalities are strongly associated with data inequalities[10]: existing urban data are disproportionately abundant on the perspectives and realities of privileged social groups and affluent urban areas. The use of these uneven data to train AI-based urban models often inadvertently embeds biases in algorithms and amplifies their social consequences[11]. These models are often formulated and parameterised without knowledge of local realities, resulting in inaccurate definitions that do not account for urban disparities, as they rely on broad-brush generalisations such as urban/rural or slum/non-slum dichotomies[12,13]. For instance, populations in urban deprived areas are under-represented in censuses of Low and Middle-Income Countries (LMICs), and current population modelling algorithms based on these data have been shown to underestimate the number of people living in slums[14,15].

Second, the social inequality challenges extend beyond data scarcity to include how data systems are framed and governed. Fragmented, siloed data systems and methodological divides between "big data" and community-based data approaches[16] often undermine coordinated, action-oriented responses and ethical engagement with residents[17]. Furthermore, the framing of problems by and through data is neither neutral nor uncontested but reflects the perspectives, worldviews, and practices of those involved in the design of data systems[18,19]. In the case of deprived urban areas, this is amplified by the fact that the very definition of a slum or informal settlement is highly variable across countries, and is often disputed even within a country, with divergent views between governments and social movements[20]. A related example is the recent decision of the Brazilian National Statistics Office to change the census terminology from "abnormal agglomerates" to "urban communities and favelas" in response to strong concerns expressed by social movements led by favela residents[21]. Without careful attention to such framing issues, data-driven and AI-powered urban models risk perpetuating epistemic injustices[22] by marginalising the voices and worldviews of those most affected.

Recent calls for transdisciplinary urban science echo concerns about potential injustices[23,24] and suggest leveraging citizen science data to inform decision-making[8]. Yet most existing studies that use AI and EO to analyse urban deprivation do not meaningfully engage with the communities they aim to represent (i.e., the "data subjects"). Instead, they often adopt extractive approaches to "citizen science"[25] that rely on limited stakeholder input, typically from more privileged groups, thereby risking misrepresentation of the highly diverse realities of marginalised urban residents. While industry-led efforts promote "data ethics" and "AI for social good", they are insufficient to address the contextual and political complexities of mapping urban deprivation across the globe and may contribute to what has been termed "data colonialism"[21].

This article presents the Integrated Deprived Area Mapping System (IDEAMAPS) Data Ecosystem: a justice-oriented framework for co-creating urban data systems, grounded

in a participatory analytics approach that integrates community knowledge and agency with data science and AI. It addresses the urgent need for systematic, participatory approaches that support the equitable and responsible use of AI in urban science and practice[24,26]. Underpinned by participatory action research, equitable co-creation processes and a co-designed digital collaboration platform, the IDEAMAPS Data Ecosystem approach aims to deliver locally meaningful data and computational models that foster transformative urban innovation.

This article makes two key contributions: (a) it offers a set of principles, processes and digital tools for co-creating justice-oriented urban data ecosystems, developed through collaborations with communities, governments, and academia; (b) based on empirical results in our pilot cities in Kenya (Nairobi) and Nigeria (Kano and Lagos), we provide evidence that our approach can produce novel, improved models for characterising morphological urban informality to effectively support policy and decision-making. Unlike many AI approaches that are purportedly participatory but primarily crowdsource labels and disregard power imbalances[27], our framework uses meaningful engagement with communities and stakeholders to shape core model definitions, surface inherent trade-offs, and evaluate the broader socio-political implications of how deprivation is represented in data and models.

2. Results

Principles and processes for co-creating Just Urban Data Ecosystems

IDEAMAPS emerged as a network of researchers, practitioners, and civic actors interested in advancing methods for understanding and improving deprived urban areas[16,28,29]. Building on these early collaborations, the IDEAMAPS Data Ecosystem project developed a participatory approach to co-creating AI-powered urban data ecosystems, guided by four principles, summarised in Box 1: grounded, transformative, transferable, and feasible.

These principles are implemented through an iterative, transdisciplinary co-creation process (Figure 1) to develop equitable and impactful data ecosystems. Our approach considers urban data ecosystems to comprise not only locally validated data and responsible urban AI models but also new action-oriented knowledge and collaborative relationships among diverse stakeholders. This four-stage co-creation cycle can be repeated several times, facilitating the gradual transformation of data into tangible impacts, as described in the Methods section.

BOX 1. Principles for Co-creating Just Urban Data Ecosystems

Grounded. *Urban data models must be rooted in the lived experiences, priorities, and worldviews of the people they aim to represent.* This requires sustained, meaningful engagement with disadvantaged groups to ensure that data outputs are accurate, up to date, and aligned with local needs and priorities. Divergent views, trade-offs, and potential harms must be explicitly discussed and addressed through inclusive dialogue.

Transformative. *Data outputs should provide action-oriented insights that enable policymakers and urban practitioners to address structural inequalities*. By informing interventions that expand access to services and infrastructure to fulfil the rights of marginalised populations, urban models can help reduce distributional injustices and foster more equitable futures.

Transferable. *Urban models should be adaptable across diverse urban contexts*. Engagement with regional and transnational networks can ensure that modelling approaches are translated to reflect local challenges, values, and governance structures, while maintaining core principles of justice and transformation.

Feasible. *Justice-oriented models must be both technically viable and accessible to their audiences*. Leveraging advances in AI and EO requires computational efficiency and the use of accessible, scalable technologies that can be implemented in resource-constrained settings and are easily interpretable by the intended decision-makers.

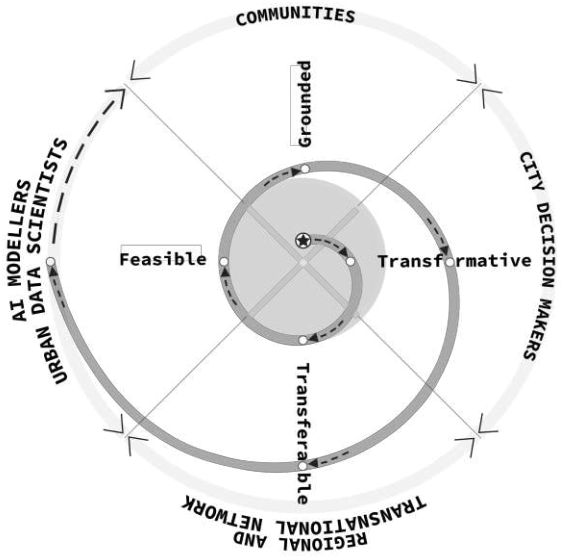


Figure 1. A justice-oriented co-creation process for AI-powered urban data ecosystems. *The process centres on structurally marginalised communities and comprises four iterative steps that can be repeated and refined over time as needed. Each step operationalises one of the core principles outlined in Box 1 (grounded, transformative, transferable, and feasible), ensuring that the resulting data ecosystem is inclusive, context-sensitive, and actionable.*

The IDEAMAPS Data Ecosystem Platform: a tool for participatory modelling and validation

To support our justice-oriented co-creation process, we implemented a web platform for data-driven citizen science that enables evolutionary participatory modelling and

validation (Figure 2). After these models are specified in collaboration with local communities, researchers, city stakeholders, and domain experts (e.g., city planners or public health professionals), the model outputs are made available on the digital platform. The platform interface presents data on a 100m x 100m grid, which not only provides a uniform way to present model outputs but also helps protect community privacy, as revealing exact boundaries could lead to potential harm (e.g., forced evictions). Each grid cell is shown in a colour corresponding to classification outputs from urban computational models, categorised into three bands: Low (Blue), Medium (Orange), or High (Red). The three-band scale makes the implications of model outputs tangible to citizens by using understandable categories rather than numeric values, which are often harder to interpret.

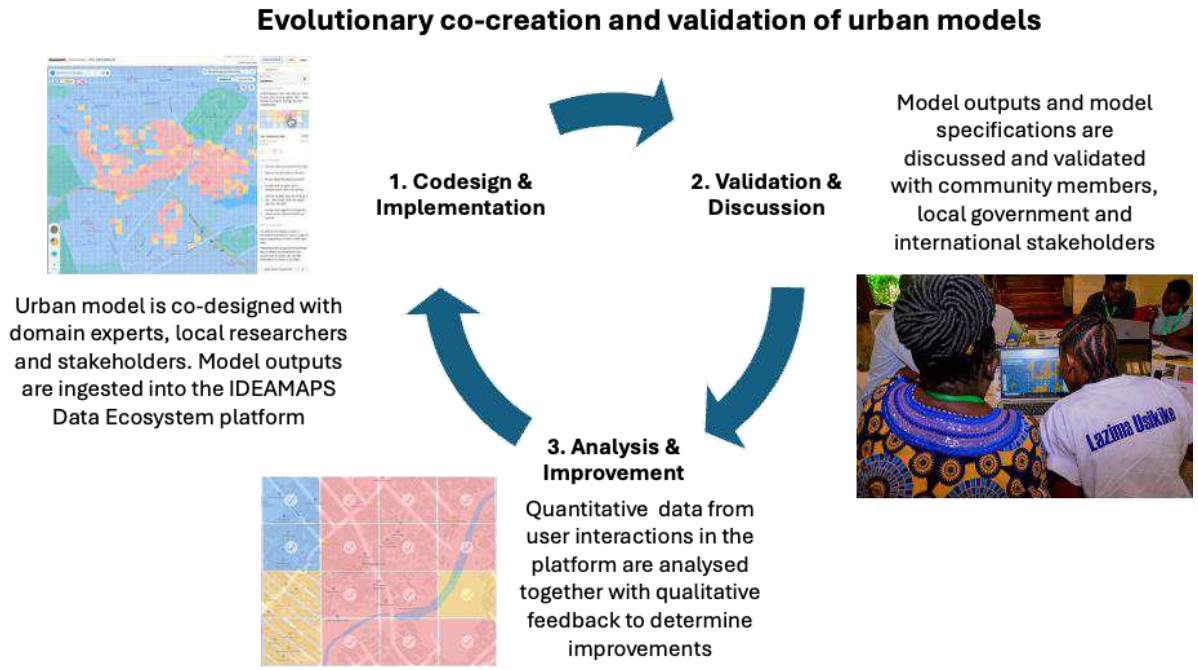


Figure 2. Participatory development, stepwise improvement, and validation of urban models.

Even when urban models use the best available data, they often fail to adequately represent local perspectives and conditions, for example, due to biases in the training datasets or incorrect assumptions. To improve this on our platform, citizens can select cells corresponding to different urban areas to agree with the presented classifications or adjust them to match their lived experience. This participatory validation is repeated with multiple social groups and city stakeholders to gather diverse views on how urban areas are represented in the model-generated data. Alongside the quantitative data collected through these interactions with the Ecosystem, we also gather qualitative feedback. This allows citizens to share their views on how their neighbourhoods are depicted and framed in the data, helping us understand the potential opportunities and threats for affected communities.

The validation data generated by citizens on the platform are analysed alongside qualitative feedback on model definitions from all groups to identify potential

improvements. Qualitative feedback helps identify potential opportunities and threats arising from the model and tests its underlying assumptions. Quantitative validation, in turn, provides modellers and urban scientists with precise data to refine the modelling approach and parameter choices for subsequent versions. These improved versions then feed into the next iteration of the process, allowing initial baseline models to be continually refined through collective scrutiny, discussion, and further enhancement.

Urban data ecosystem: a multidimensional characterisation of urban informality

The IDEAMAPS Data Ecosystem digital platform and participatory validation process have been applied through various engagements in our pilot cities: Kano, Lagos and Nairobi, enabled by participatory action research groups with partner communities involving cross-sectoral multi-stakeholder workshops and participatory modelling and software co-design (details in Methods). Based on a systematic list of potential Domains of Deprivation[28], these engagements identified actionable priorities for urban data models that could support urban regeneration.

Table 1 presents a description of the models we developed and made available on our digital platform. Together with stakeholder interactions, the suite of models for each city constitutes a data ecosystem that supports the characterisation of urban areas across several domains of deprivation: unplanned urbanisation (morphological informality), infrastructure (road access deprivation), access to facilities and services (general and emergency obstetric healthcare), contamination (accumulated waste piles), and hazards (flood hazard). These domains are complementary yet closely related, offering a multidimensional perspective on urban spaces.

Table 1. List of urban data models developed and available in the IDEAMAPS Data Ecosystem platform.

Urban model *(Domain of Deprivation)*	**Description of the model**
Morphological Informality *(Unplanned urbanisation)*[30–32]	Captures the contrast between planned and unplanned parts of the city, serving as a physical proxy for urban deprivation, with higher morphological informality typically found in more deprived areas.
Road Access Deprivation (*Infrastructure)*	Represents the accessibility of motorable roads from buildings and their quality in terms of surface type (paved/unpaved) as an indicator of urban deprivation.
General Healthcare Access Deprivation *(Facilities & Services)*	Indicates how difficult it is for people to access healthcare, considering availability, proximity, and types of nearby healthcare facilities.
Emergency Obstetric Care (EmOC) Access Deprivation *(Facilities & Services)*	Depicts how difficult it is for women to access emergency maternal care, considering the availability, proximity, and types of healthcare facilities that provide comprehensive EmOC services.

Accumulated Waste Piles *(Contamination)*	Shows the chance of finding piles of waste in different places. Waste piles can indicate areas that lack waste collection services.
Flood Hazard *(Hazard)*	Represents the estimated flood extent and flood depth of past extreme rainfall events in the city.

To illustrate the value added by our participatory modelling process, we detail the stepwise development and improvement of the morphological informality model, synthesised in Table 2. One of the key priorities identified by government agencies was to locate informal settlements and slums, a challenge that reflects the absence of a unified spatial definition and methodology for their recording. Our initial attempt (which we call "v0") included an "overall deprivation" model that classified each grid cell in an urban area as "deprived" or "non-deprived". We generated model outputs using sample training datasets (see details in Methods) and then brought these outputs into a first prototype of the IDEAMAPS Data Ecosystem Platform to collect quantitative validation data (i.e., how many people agreed or disagreed with the model's binary output) alongside qualitative observations shared in our workshops and focus groups with communities, local researchers, and officials. After analysing these initial validation results, several stakeholders expressed different views on what would characterise an urban area as "deprived" or "slum". This ambiguity and failure to reach consensus made it very challenging to properly validate the model outputs, as different people relied on different concepts to indicate whether they agreed with the model.

We concluded that not only did the model outputs require improvement, but the conceptual model definition itself also needed revision. In discussions with international organisations, we were referred to the "degree of urbanisation" methodology endorsed by the UN Statistical Commission[33] to provide a harmonised definition of urban areas and to avoid the ambiguous and inconsistent definitions of cities used worldwide. Inspired by this, we co-designed a more precise three-band conceptual model of the "degree of morphological informality", based on the layout of buildings and roads in neighbourhoods, as visualised in Figure 3. This definition is narrower than usual definitions of "slum" or informal settlement – but precisely because of this, it can offer a more consistent and evidence-based baseline upon which other dimensions of deprivation can be mapped and understood.

The conceptual model of morphological informality is based on two key characteristics of unplanned urban settlements: irregular settlement layout and the presence of small, dense structures (Figure 3). Settlement layout irregularity is characterised by haphazardly positioned buildings with varying distances and orientations relative to neighbouring structures. Small, dense structures, on the other hand, are characterised by closely spaced buildings, typical of highly populated, low-rise areas. These characteristics were identified as relevant to mapping morphological informality with respect to desirability and feasibility in stakeholder engagements. Levels of morphological informality are defined as "low" when the settlement layout is regular and small, and the built structures are not dense. The "medium" band indicates that

only one of the characteristics is informal, the other is formal; whereas the "high" band is used to classify areas where both characteristics are present.

Table 2. Stages of model development for the morphological informality domain, following our iterative participatory modelling approach.

Model Version	V0: overall deprivation	V1: morphological informality (baseline)	V2: morphological informality (improved)
Features and changes	The initial model attempted to identify deprived urban areas (such as slums and informal settlements) in the city. Each cell is classified as deprived/non-deprived.	The model definition was revised based on feedback. The new model on "morphological informality" distinguishes three bands (high/medium/low) depending on urban form and layout features (see Table 3).	Based on the validation of v1, a new version of the model implementation was developed, but maintaining the same three-band conceptual model. New thresholds and criteria for distinguishing high/medium/low bands were established for the semi-supervised clustering algorithm.
Engagement and feedback received	Cross-sectoral workshops and PARs with partner communities revealed that a binary classification ("deprived/non-deprived") is ambiguous, and perceptions and assessments vary widely.	Validation workshops were conducted with cross-sectoral stakeholder groups and partner communities. Quantitative and qualitative results indicated that the model concept was clear, but the classification criteria required adjustments (see Table 4).	A city-wide validation campaign was conducted with members of several informal communities in our partner cities, as well as with a wider group of cross-sectoral stakeholders. Results showed significant improvement in model performance and confirmed broad acceptance of the conceptual model.

Band	Description	Visual representation

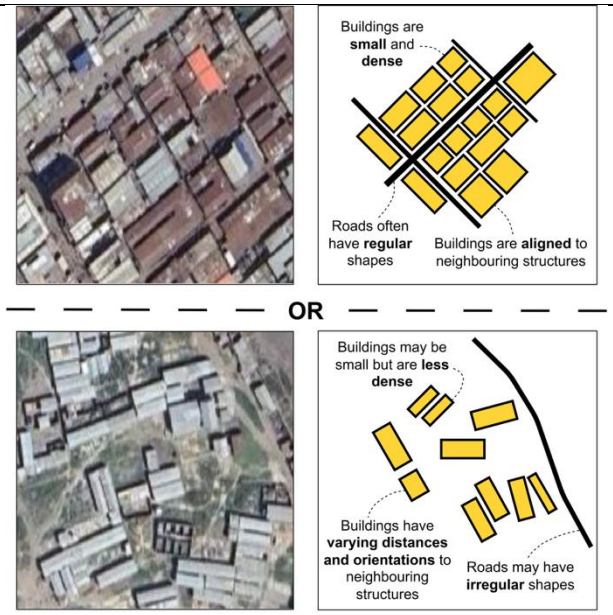
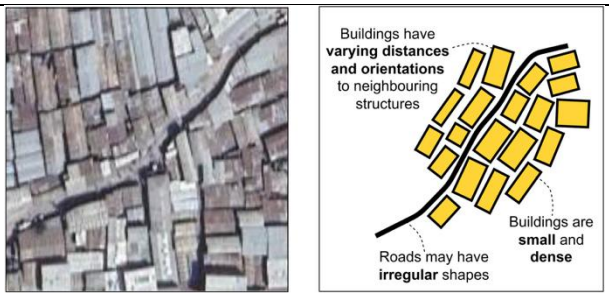
Low	Buildings are larger, detached from one another, planned and regularly arranged. It is easy to visualise roads, pathways, social amenities, or open spaces between them. Areas with no buildings are also categorised as low.	
Medium	Buildings are small and close together (it can be hard to visualise roads, pathways, or spaces between them) **or** buildings are irregularly arranged (with varying alignment to neighbouring structures). Roads are moderately present.	
High	Buildings are small, close together, and irregularly arranged (with varying alignment to neighbouring structures). Roads are hard to visualise, or they are absent.	

Figure 3. A conceptual model for the degree of morphological informality of urban areas. *The model classifies urban areas into three deprivation bands (high/medium/low) based on the degree of informality of their urban form.*

Following our iterative co-creation process, this model (hereinafter, "v1" baseline model) underwent a second validation interaction loop with stakeholders, beginning with a re-grounding of the model (Figure 1). As part of this process, we conducted validation workshops with local communities to assess the baseline model's outputs using the IDEAMAPS Data Ecosystem Platform. Quantitative citizen data and qualitative discussions with communities revealed that the conceptual model was clearer and

better understood, but also that there was room to improve classification accuracy. Figure 4 visualises the differences between model outputs and user validations.

To improve the model in the second iteration of the co-creation process, we considered the validations and discussions during the PAR sessions. Since community members agreed with the conceptual model but indicated that the baseline model needed improvement, we kept the conceptual model unchanged and focused modelling efforts on a new implementation (hereinafter, "v2" or the improved model) by changing the algorithm used to identify small, dense structures and irregular layouts. Subsequently, we conducted a more comprehensive citywide validation campaign to collect data from a wider range of communities and stakeholders (e.g., government officials, local academics) across our pilot cities (see Methods).

Table 3 presents an analysis of the citizen validation data we collected, comparing model accuracy using F1 scores. It shows that the improved model implementation (v2) better reflected stakeholders' perspectives, improving overall accuracy for Lagos (40%) and Kano (37%), while for Nairobi, the overall accuracy remained almost the same. Nevertheless, model v2 shows a strongly improved classification for areas with medium levels of morphological informality, with F1 scores jumping up by 61-450%.

The improvements are visually clear in the maps comparing different model versions for one of our partner communities in Korogocho, Nairobi (Figure 5): the baseline model represents morphological informality as predominantly low (formal) or high (informal). In contrast, community member validations indicate that urban form is often a mixture of formal and informal elements, represented at the medium level; this is better captured in our v2 model. This finding highlights the limitation of dichotomous deprivation modelling and underscores the importance of considering mixed forms. The improved model outputs offer a more nuanced view of morphological informality, which, although independently defined, aligns more closely with the community-sourced validation data for most categories in our three pilot cities (Table 3).

Table 3. Model performance comparing results from the validation of stakeholders based on the morphological informality model v1 versus v2.

City	Model iteration	Acc (%)	F1 score		
			Low	Medium	High
Nairobi	Baseline (V1)	66.0	0.801	0.104	0.490
	Improved (V2)	67.5	0.743	0.574	0.681
	Rel diff	+ 2 %	- 7 %	+ 450 %	+ 39 %
Lagos	Baseline (V1)	41.8	0.523	0.384	0.184
	Improved (V2)	58.5	0.542	0.644	0.549
	Rel diff	+ 40 %	+ 4 %	+ 68 %	+ 198 %
Kano	Baseline (V1)	47.3	0.601	0.391	0.214
	Improved (V2)	64.7	0.708	0.629	0.521
	Rel diff	+ 37 %	+ 18 %	+ 61 %	+ 144 %



Figure 4. Alluvial plots comparing, for each city, the model-predicted deprivation level (high/medium/low, left-hand side) with the corresponding user validations (right-hand side). The width of each ribbon represents the number of spatial units in that prediction–validation combination: horizontal ribbons indicate agreement, while diagonal ribbons show where the model over- or under-estimated deprivation relative to user assessments.

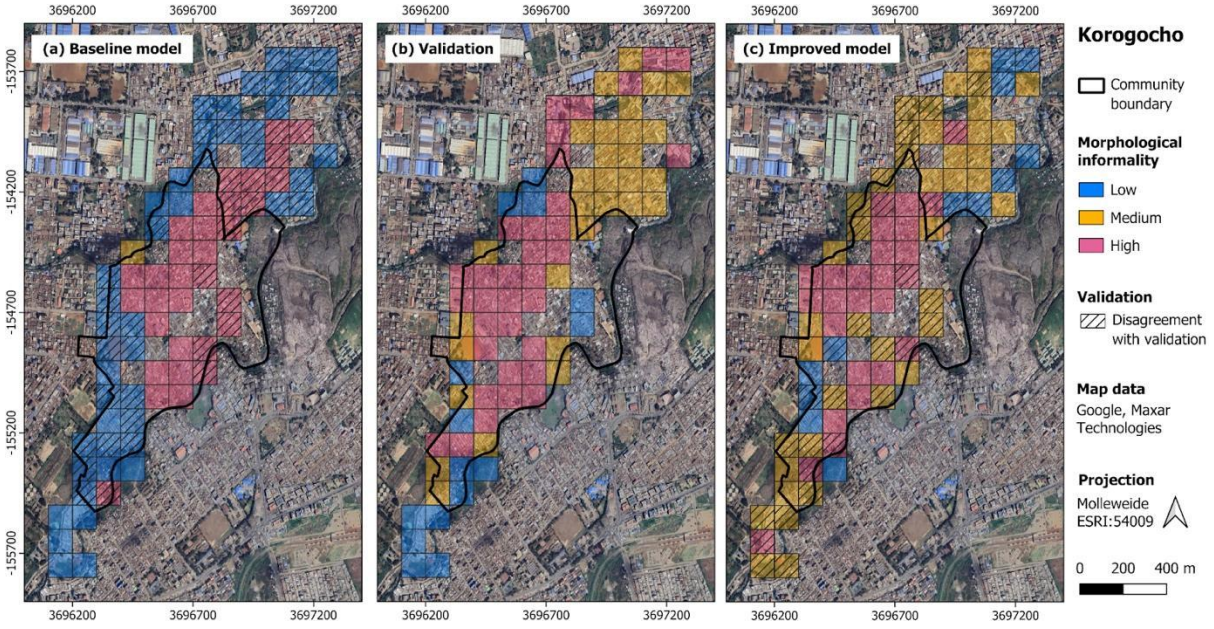


Figure 5. Comparison between model results and the data from model validation. Data for our partner community in Korogocho, Nairobi, Kenya (top): (a) initial baseline model outputs; (b) community validation data; (c) improved model after learnings from the validation and discussion sessions.

3. Discussion

The participatory approach of our IDEAMAPS Data Ecosystem framework has proven useful for generating highly accurate, socially validated urban data to address the invisibility of informal settlements and disadvantaged urban communities. We discuss three main contributions from this work for urban research and practice.

Grounding urban models on community assets and priorities

A distinctive feature of the modelling approach presented here is its grounding in in-depth engagement with partner communities across all cities through participatory action research and transdisciplinary co-design. The decisions regarding which models to include for each city were driven by these interactions and by data collected by communities. This approach builds on and reinforces the longstanding and important work of community-led data generation by grassroots organisations, including members of the Slum Dwellers Federations[34,35]. It ensures that the data reflect community priorities and perspectives while mitigating risks associated with community visibility in the data, in line with recent recommendations in the literature[16,17].

The feedback from our partner community members emphasised the value of the co-creation and modelling process in developing a critical understanding of how available data and new urban models portray their communities. This critical consciousness can itself be a pathway to transforming perceptions and behaviour[36]. To "ground" the participatory modelling process, we first worked with communities to analyse existing maps, then undertook participatory mapping to fill the gaps they identified in current cartographic representations (see details below in Methods).

During this process, our team listened carefully to community concerns about the stigmatisation of neighbourhoods when data focuses only on deprivation, casting them as a "burden to be removed" and potentially justifying top-down evictions. In response, we adopted a new approach that begins by mapping local community assets, i.e. points of reference selected by community members themselves (e.g., street vendors, water points, community centres). Curated into OpenStreetMap (OSM), these assets are now widely accessible and included as a base layer in our validation tool. This not only provides reference points for communities to interpret the maps but also highlights their significant contributions to city life and economies, underscoring that they are not the problem but part of the solution in co-creating just and liveable urban spaces. In our feedback workshops, community members expressed pride in seeing their neighbourhoods mapped and noted positive side effects, such as identifying most local facilities and being able to request car-riding services, which can now reach their communities thanks to the data created in OSM.

A more precise and multidimensional characterisation of urban deprived areas and their degree of informality

Our discussions with cross-sectoral stakeholders showed that efforts to develop a precise methodology for mapping and measuring slum areas—currently lacking in SDG 11 indicators—are hindered by limited conceptual clarity and ambiguous understandings of what constitutes a "slum" or informal settlement. This challenge is amplified by the diverse forms these urban deprived areas take globally[37] and the many proposed features and dimensions of deprivation used to describe them[28].

To overcome this impasse, we depart from usual approaches to mapping slums using Earth Observation and AI[38], which are commonly based on binary classification models

of slum/non-slum areas, in two important ways. First, our IDEAMAPS Data ecosystem approach consists of a suite of data models and maps across distinct domains that characterise urban areas using a uniform three-band classification system with well-defined, objective criteria for the high/medium/low classes. These maps thus reveal multiple dimensions of the challenges and injustices faced by residents of deprived urban areas, which can be objectively verified and addressed.

Second, our results help to consolidate a definition of urban informality based on morphological criteria, in line with recent literature[39–42]. In contrast to previous approaches, which relied on definitions established solely by remote sensing experts, our conceptual model of morphological informality has been extensively discussed with, and broadly accepted by, stakeholders. Our findings show that this urban informality model can serve as an effective baseline layer, helping to explain the distribution of areas that emerge from organic urban growth. In addition, the data models improved significantly through our interactive validation process, now offering a more nuanced representation of urban areas that aligns more closely with stakeholders' perspectives, particularly in areas with high and medium levels of morphological informality. With F1 improvement rates ranging from 39% to an impressive 450%, the enhanced data can substantially support the identification of priority areas for urban upgrading.

Nevertheless, our improved models still did not fully match stakeholder classifications, showing that the polysemic nature of urban deprivation cannot be fully captured by morphological informality. The remaining mismatches may be due to several factors, ranging from outdated data underlying urban models (e.g., incomplete building footprints and street networks) to residual ambiguity in the definition of informality. Nevertheless, we found that the value of our validation process lies as much in showing convergence as in making explicit the reasons for divergence: urban data models in our approach provide a baseline for discussion, which stakeholders have found useful even when they were not perfect.

Co-creation of living urban data ecosystems as transformative urban regeneration

The IDEAMAPS data ecosystem approach reconceptualises urban modelling as part of an interactive process for co-creating urban knowledge and data. It is grounded in transdisciplinary interactions among communities, city government officials, policymakers at different levels, and urban data scientists and researchers. Our study shows that co-creating urban models with affected communities is not only an ethical imperative but also an effective means of making urban science more accurate and transformative. Our interactive process makes three key contributions: (a) quantitative assessment of model classification accuracy using citizen-generated data; (b) qualitative evaluation of the efficacy and interpretability of underlying model definitions; and (c) improvement of the model's capacity to support action and decision-making. In addition to the models co-created in our pilot cities in Kenya and Nigeria presented here, we have generated beta models for additional cities to engage a broader audience, including cities in Latin America (Colombia) and other African countries (Ghana, Mozambique).

Our approach envisions a living, open ecosystem that can grow and learn over time by incorporating new perspectives and generating new knowledge and citizen data with each iteration. The IDEAMAPS validation tool is open and can be easily fed with third-party models and datasets beyond those we created. This enables researchers to move beyond narrow, tokenistic approaches to "participatory" AI that focus solely on collecting new labels for model training and rely on shallow notions of "empowerment" that disregard power asymmetries[27], agency, and the distinctive epistemologies and ontologies of the mapped communities[25]. The participatory analytics approach presented here charts a path for future urban science that embeds systematic mechanisms to ensure political control and the governance of the purposeful use of AI-enabled models, thereby effectively addressing the urgent task of improving the living conditions, health, and well-being of the most impoverished and disadvantaged urban citizens.

4. Methods

The IDEAMAPS Data Ecosystem project reported in this article employed a mixed-methods approach that combines three streams: (a) participatory action research groups with partner communities; (b) cross-sectoral co-creation workshops in our three pilot cities - Lagos, Kano (Nigeria), and Nairobi (Kenya); (c) participatory software co-design and modelling; and. In the next section, we briefly describe how these three streams are articulated to operationalise the multi-step co-creation process shown in Figure 1, and the subsequent sections explain each stream in detail.

4.1 Overview of the methodological design

The loop in Figure 1 begins with community-based participatory action research (PAR), which centres the data ecosystem on the lived experiences and priorities of residents in deprived urban areas. Drawing on critical pedagogical approaches[25], this substantial engagement helps avoid tokenism and ensures the ecosystem is firmly grounded in the perspectives and worldviews of those most affected. Implementing this step is inherently complex, particularly in heterogeneous and rapidly changing urban neighbourhoods. Multiple perspectives may coexist within and across communities, requiring facilitators attuned to local dynamics and sensitive to how intersecting social factors (such as ethnicity, occupation, age, and gender) shape what is considered valuable or visible in the urban landscape.

In the second step, insights from community engagement are incorporated into city-level, cross-sectoral co-creation workshops that involve relevant decision-makers, such as urban planners, health practitioners, academics, and civic leaders. These workshops build on the assets and priorities identified by communities to identify shared priorities for citywide urban data models that can support actionable interventions and thereby enable transformative change. The third step in our process extends this dialogue to regional and transnational networks (e.g., slum dwellers federations, NGOs, and development agencies) to explore how the models could be transferred to and adapted to other contexts. In the final step of the inner loop, AI

modellers and data scientists respond to the knowledge and priorities surfaced in the previous steps by developing feasible models and digital platforms. These tools are designed not only for technical robustness but also for accessibility, interpretability, and usability by diverse stakeholders.

Our process recognises that tensions and trade-offs may arise as new knowledge is co-created. For example, priorities identified by communities or local governments may shift when translated into transferable computational algorithms. To address this, the process anticipates a second iteration (represented as the outer loop in Figure 1), which begins by re-grounding the work with community groups. These groups are invited to scrutinise and assess the outputs of the first loop, identifying both potential risks to mitigate and opportunities for further exploration of the data to support transformative action.

The revised insights are then reintroduced into multi-stakeholder workshops to refine transformative pathways informed by the newly co-produced evidence. The final two steps of the outer loop revisit transferability and feasibility: regional and global partners assess how the approach might be replicated in other contexts, while modellers and developers focus on scaling up the urban models. These steps ensure that city-level results can be transferred to other contexts. They explore new opportunities and mechanisms for replicating the approach through regional and transnational networks, and modellers and developers work to scale the resulting AI models and platforms. Throughout the process, systematic documentation and reflection captured diverse perspectives and recorded key decisions and lessons learned, ensuring transparency and accountability.

4.2 Stakeholder engagement and co-creation

Following the co-creation approach previously outlined, our mixed-methods study design includes a series of engagements in our pilot cities of Kano, Lagos, and Nairobi from February 2023 to September 2025. Table S2 in Supplementary Information presents an overview of the engagement events, participants, and methods and activities used. The engagement events included two main types: (a) cross-sectoral engagement in multistakeholder events; (b) participatory action research (PAR)[43] activities with two partner communities in each city: Rimin Kebe and Dorayi Karama in Kano; Ajegunle-Ikorodu and Okerube-Alimosho in Lagos; Korogocho and Viwandani in Nairobi.

In addition to these two stakeholder engagement methods, our co-creation strategy involved participatory software co-design and modelling (described below). To support this process, we appointed a city liaison in each of our study cities, who was currently employed by the local government in departments interested in our study. Our government liaison partners were members of our team and attended all project-planning and co-design meetings. This study followed strict ethical protocols approved by institutional and national review boards in Kenya, Nigeria, and the UK.

The next sections provide details about these methods.

Cross-sectoral workshops

We organised three sets of cross-sectoral workshops in each of our study cities: (a) *inception workshops*: presenting project goals and methods and inviting stakeholders to provide feedback and propose ideas or initiatives that could be aligned with our study; (b) mid-term workshops: presentation of intermediary findings, including preliminary model results, discussion of next steps; (c) closing events: presentation of final results and discussion of pathways to use the study outputs in practice.

Each workshop included representatives from local and national governments, local and international NGOs, local academics/researchers, private-sector organisations, and community-based organisations. To identify participants, local teams, supported by our government liaison partners, created initial lists with people from these various profiles, building upon existing relationships, who were invited and asked to suggest further names based on their relevance to the project in a snowball approach.

In addition to these targeted invitations, the workshops were open: we disseminated an open call for participation and accepted registrations on-site on the day of the workshop. Each workshop followed a similar structure: initial guest presentations from governmental agencies and community leaders, followed by our research team's presentations, and structured feedback through focus group discussions. A full overview of all completed cross-sectoral workshops and four PAR cycles is provided in Supplementary Information, including participant numbers and profiles.

Following the completion of the major phases of our work, including the main cross-sectoral workshops, we organised learning workshops with our transdisciplinary team to promote cross-learning among teams in different cities.

Participatory Action Research (PAR) groups

The PAR sessions were organised in cycles that involved planning, data collection, collaborative analysis and reflection. Initial co-planning considered priorities identified by communities and activities proposed by the research team. After activities were jointly defined and planned, we conducted capacity-strengthening sessions on the tools used (e.g., KOBOToolBox, OpenStreetMap), and the mapping and data generation activities were facilitated. The results were then presented and jointly analysed. Participant selection was based on agreed profiles to ensure representativeness of all groups. The participants were selected based on age, sex, religion, and disability status, ensuring representation of at least one individual from each category. A core group of PAR co-researchers from the community has been consistent across various PAR activities, enabling the project to build a sustained relationship with the partner communities.

Participatory software co-design and modelling

Alongside our external stakeholder engagement, we adopted a user-centred, participatory software design methodology to develop our IDEAMAPS Data Ecosystem tool. Building on previous work[45], our method was based on principles of agile software development and incorporated a critical pedagogical lens through iterative cycles of four steps:
1) *Establishing foundations*: co-creating a vision for the software tools with the team members and discussing this vision with external stakeholders in inception meetings.
2) *Iterative software development sprints:* generation of designs and usable software tools by the research software engineering team, employing a rapid prototype approach, following a four-week "sprint" cycle.
3) *Transdisciplinary team co-design sessions*: design concepts and prototypes were extensively discussed and tested among our transdisciplinary international team (including government liaison partners). These sessions were held at least monthly, and in some crucial moments of the project, biweekly.
4) *Evaluation with external end-users*: after internal validation and corresponding improvements, the resulting software tool and outputs were discussed in our external engagement events (Table 2) with our partner communities in PAR sessions and then brought to wider discussion with multiple stakeholders in cross-sectoral workshops.

The iteration of these steps produced the IDEAMAPS Data Ecosystem validation app, presented in the results section. Alongside co-designing the software, we also run co-design sessions to develop the morphological informality model, following the process previously described (Figure 2). Full technical details of the first version of the model (v1) are reported elsewhere[46], with the focus of the present article lying on the co-design and validation process. Here, we summarise the main features of the model and the changes introduced in the improved version (v2), previously reported in this article.

The morphological informality model operationalises our co-designed conceptual model (Figure 3) through indicators derived from open building datasets. To do this, we first grouped the city's grid cells into clusters of similar urban form types and identified clusters characterised by irregular settlement layouts or small, dense structures. Specifically, we generated three urban form elements, i.e., buildings, plots, and blocks, from open global building footprints and road data obtained from the Open Buildings dataset provided by Google[47].

Urban morphometrics[48] were then employed to develop two sets of urban form characteristics for each sub-domain: irregular layout and small, dense structures. These features quantitatively describe the morphology of urban elements. To calculate the metrics, we used the Python packages OSMnx and *momepy*, which are part of PySAL (Python Spatial Analysis Library) and are built on top of GeoPandas. The characteristics were aggregated at the grid level (100 x 100 m), and the k-means clustering algorithm was applied to group grid cells with similar urban forms. We then worked with local team members and stakeholders to interpret these clusters and determine the sub-domains they represent.

For the improved model ("v2"), we kept the same conceptual framework but refined the implementation. We observed that sharing urban form characteristics across sub-

domains resulted in strong correlations, particularly between areas with irregular settlement patterns and small, dense structures. To lower these correlations, we removed shared urban-form features. Additionally, we added new characters to better represent the irregular settlement layout sub-domain.

An overview of the validation data used in this article is provided in the Supplementary Information. Validation data were collected in Nairobi with community members from Korogocho and Viwandani on 4 and 5 June 2025, respectively. The city-wide validation campaign in Nairobi was conducted on 30 July 2025 (validations on 1 August 2025 were also included). Validation data in Kano were collected for the partner communities Rimin Kebe and Dorayi Karama on 22 May 2025, and the city-wide validation campaign was conducted from 27 June to 3 July 2025. In Lagos, the city-wide validation campaign, including members of the local partner communities Okerube and Ajegunle-Ikorodu, was conducted on 11 and 12 June 2025.

To analyse the quantitative feedback from the validation tool, model outputs are evaluated using the metrics F1 score and overall accuracy (Acc), defined as follows:

$$F1 = \frac{TP}{TP + \frac{1}{2} \times (FP + FN)}$$

$$Acc = \frac{TP + FN}{TP + TN + FP + FN}$$

Where *TP, TN, FP,* and *FN* denote the number of true positives, true negatives, false positives, and false negatives, respectively.

Qualitative feedback from PAR sessions and cross-sectoral workshops was recorded in audio transcripts and detailed notes taken by the local research team. We undertook inductive thematic analysis to address the research questions, with initial analysis conducted by each city team and collaborative analysis conducted with the entire research team in cross-learning workshops.

Data and Code Availability

All model outputs used to generate the tables and charts in this article are openly available at Zenodo, alongside the data from all other models generated by the IDEAMAPS Data Ecosystem project (Table):
https://zenodo.org/communities/ideamaps/ (Accessed 19 January 2025).
The documentation, source code and outputs for models developed within the IDEAMAPS Data Ecosystem project that address different domains of deprivation at the GitHub repository:
https://github.com/urbanbigdatacentre/ideamaps-models/tree/dev (Accessed 19 January 2025).

Acknowledgments

This article was generated as part of the IDEAMAPS Data Ecosystem project and the authors acknowledge the contributions from the stakeholders and partners communities in Lagos (Nigeria), Kano (Nigeria) and Nairobi (Kenya), as well as from the whole team including: João Porto de Albuquerque, Bunmi Alugbin, Kehinde Baruwa, Andrew Mitham-Clarke, Peter Elias, Helen Elsey, Ryan Engstrom, Grace Gielink, Serkan Girgin, Diego Pajarito Grajales, Angela Abascal, Caroline Kabaria, Monika Kuffer, Sebastian Hafner, Oluwatoyin Odulana, Francis Onyambu, Oluwatimilehin Adenike Shonowo, Dana Thomson, Grant Tregonning, Qunshan Zhao, Mingyu Zhu, Xingyi Du, Sabitu Da'u Sa'adu, Taiwo Gboluwaga Amole, Esther Judith, Caleb Cheruiyot, Darren Dixon, Mansur Mahmoud Ahmad.

This work was supported, in whole or in part, by the Bill & Melinda Gates Foundation INV-045252. Under the grant conditions of the Foundation, a Creative Commons Attribution 4.0 Generic License has already been assigned to the Author Accepted Manuscript version that might arise from this submission. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation.

Author Contribution Statement

JPA: Conceptualization; Methodology, Writing – Original Draft, Supervision, Project Administration, Funding Acquisition
SH: Software, Visualisation
PE, CK: Project Administration, Supervision
All authors: Investigation; Writing - Review & Editing

References

1. Hosny, A. & Aerts, H. J. W. L. Artificial intelligence for global health. *Science* **366**, 955–956 (2019).
2. Abascal, A. *et al*. AI perceives like a local: predicting citizen deprivation perception using satellite imagery. *Npj Urban Sustain*. **4**, 1–14 (2024).
3. Federspiel, F., Mitchell, R., Asokan, A., Umana, C. & McCoy, D. Threats by artificial intelligence to human health and human existence. *BMJ Glob. Health* **8**, e010435 (2023).
4. Kuffer, M. *et al*. The role of earth observation in an integrated deprived area mapping 'system' for low-to-middle income countries. *Remote Sens*. **12**, (2020).

5. Burke, M., Driscoll, A., Lobell, D. B. & Ermon, S. Using satellite imagery to understand and promote sustainable development. *Science* **371**, eabe8628 (2021).
6. *Rescuing SDG 11 for a Resilent Urban Planet - SDG 11 Synthesis Report High Level Political Forum 2023*. (2023).
7. Ezeh, A. *et al.* The history, geography, and sociology of slums and the health problems of people who live in slums. *The Lancet* **389**, 547–558 (2017).
8. Min, Y., Chen, H. & Perucci, F. Data on SDGs are riddled with gaps. Citizens can help. *Nature* **633**, 279–281 (2024).
9. Nations, U. Transforming Our World: The 2030 Agenda for Sustainable Development - A/RES/70/1. *U. N. Gen. Assem. - 70th Sess.* **16301**, 1–35 (2015).
10. Porto de Albuquerque, J. *et al.* Dialogic data innovations for sustainability transformations and flood resilience: The case for waterproofing data. *Glob. Environ. Change* **82**, 102730–102730 (2023).
11. Shaffer, J., Alenichev, A. & Faure, M. C. The Gates Foundation's new AI initiative: attempting to leapfrog global health inequalities? *BMJ Glob. Health* **8**, e013874 (2023).
12. Dong, L. *et al.* Defining a city — delineating urban areas using cell-phone data. *Nat. Cities* **1**, 117–125 (2024).
13. Lilford, R., Taiwo, O. J. & de Albuquerque, J. P. Characterisation of urban spaces from space: going beyond the urban versus rural dichotomy. *Lancet Public Health* **3**, e61–e62 (2018).
14. Abascal, A. *et al.* Making Urban Slum Population Visible: Citizens and Satellites to Reinforce Slum Censuses. in *Urban Inequalities from Space: Earth Observation Applications in the Majority World* (eds Kuffer, M. & Georganos, S.) 287–302

(Springer International Publishing, Cham, 2024). doi:10.1007/978-3-031-49183-2_14.

15. Thomson, D. R. *et al.* Improving the accuracy of gridded population estimates in cities and slums to monitor SDG 11: Evidence from a simulation study in Namibia. *Land Use Policy* **123**, 106392 (2022).
16. Thomson, D. R. *et al.* Need for an integrated deprived area 'slum' mapping system (IDEAMAPS) in low-and middle-income countries (LMICS). *Soc. Sci.* **9**, (2020).
17. Borofsky, Y. *et al.* An agenda for data-rich, action-oriented, ethical research on infrastructure in informal settlements. *Nat. Cities* **2**, 1026–1036 (2025).
18. Dencik, L. & Sanchez-Monedero, J. Data justice. *Internet Policy Rev.* **11**, 0–16 (2022).
19. D'Ignazio, C. & Klein, L. F. *Data Feminism*. (The MIT Press, Cambridge, Massachusetts, 2020).
20. Pratomo, J., Kuffer, M., Martinez, J. & Kohli, D. Coupling Uncertainties with Accuracy Assessment in Object-Based Slum Detections, Case Study: Jakarta, Indonesia. *Remote Sens.* **9**, 1164 (2017).
21. *Nota Metodológica No. 1: Sobre a Mudança de Aglomerados Subnormais Para Favelas e Comunidades Urbanas*. https://agenciadenoticias.ibge.gov.br/media/com_mediaibge/arquivos/4823879b95c3b36b71b8b11f6a143d45.pdf.
22. Bhakuni, H. & Abimbola, S. Epistemic injustice in academic global health. *Lancet Glob. Health* **9**, e1465–e1470 (2021).
23. Lobo, J. *et al.* Integration of urban science and urban climate adaptation research: opportunities to advance climate action. *Npj Urban Sustain.* **3**, 1–9 (2023).

24. Acuto, M., Parnell, S. & Seto, K. C. Building a global urban science. *Nat. Sustain.* **1**, 2–4 (2018).
25. de Albuquerque, J. P. & Almeida, A. A. de. Modes of engagement: reframing 'sensing' and data generation in citizen science for empowering relationships. in *Toxic Truths: Environmental Justice and Citizen Science in a Post Truth Age* nn-nn (Manchester University Press, Manchester, UK, 2017).
26. Chambers, J. M. *et al.* Co-productive agility and four collaborative pathways to sustainability transformations. *Glob. Environ. Change* **72**, 102422–102422 (2022).
27. Maas, J. & Inglés, A. M. Beyond Participatory AI. *Proc. AAAIACM Conf. AI Ethics Soc.* **7**, 932–942 (2024).
28. Abascal, Á. *et al.* 'Domains of Deprivation Framework' for Mapping Slums, Informal Settlements, and Other Deprived Areas in LMICs to Improve Urban Planning and Policy: A Scoping Review. *Preprints* **93**, (2021).
29. Elias, P. & de Albuquerque, J. P. Data and the Localization of Sustainable Development Goals in Africa: The Case of SDG 11 in Lagos and Accra. in *Localizing the SDGs in African Cities* (eds Croese, S. & Parnell, S.) 115–131 (Cham, Switzerland, 2022). doi:10.1007/978-3-030-95979-1_8.
30. Hafner, S. *et al.* Morphological Informality Modelled Dataset in Nairobi, Kenya. Zenodo https://doi.org/10.5281/ZENODO.17963362 (2026).
31. Hafner, S. *et al.* Morphological Informality Modelled Dataset in Lagos, Nigeria. Zenodo https://doi.org/10.5281/ZENODO.17963614 (2026).
32. Hafner, S. *et al.* Morphological Informality Modelled Dataset in Kano, Nigeria. Zenodo https://doi.org/10.5281/ZENODO.17965823 (2026).

33. European Commission. Statistical Office of the European Union. *Applying the Degree of Urbanisation: A Methodological Manual to Define Cities, Towns and Rural Areas for International Comparisons : 2021 Edition*. (Publications Office, LU, 2021).

34. Patel, S. & Baptist, C. Editorial: Documenting by the undocumented. *Environ. Urban.* **24**, 3–12 (2012).

35. Satterthwaite, D. *et al.* Building Resilience to Climate Change in Informal Settlements. *One Earth* **2**, 143–156 (2020).

36. Porto de Albuquerque, J. *et al.* The role of data in transformations to sustainability: a critical research agenda. *Curr. Opin. Environ. Sustain.* **49**, 153–163 (2021).

37. Yeboah, G. *et al.* Analysis of OpenStreetMap Data Quality at Different Stages of a Participatory Mapping Process: Evidence from Slums in Africa and Asia. *ISPRS Int. J. Geo-Inf.* **10**, 265–265 (2021).

38. Kuffer, M., Pfeffer, K. & Sliuzas, R. Slums from Space—15 Years of Slum Mapping Using Remote Sensing. *Remote Sens.* **8**, 455–455 (2016).

39. Friesen, J., Taubenböck, H., Wurm, M. & Pelz, P. F. Size distributions of slums across the globe using different data and classification methods. *Eur. J. Remote Sens.* **52**, 99–111 (2019).

40. Wurm, M. & Taubenböck, H. Detecting social groups from space – Assessment of remote sensing-based mapped morphological slums using income data. *Remote Sens. Lett.* **9**, 41–50 (2018).

41. Bettencourt, L. M. A. & Marchio, N. Infrastructure deficits and informal settlements in sub-Saharan Africa. *Nature* **645**, 399–406 (2025).

42. Li, C. *et al.* Mapping urban slums and their inequality in sub-Saharan Africa. *Nat. Cities* 1–12 (2025) doi:10.1038/s44284-025-00276-0.

43. Cornish, F. *et al.* Participatory action research. *Nat. Rev. Methods Primer* **3**, 1–14 (2023).

44. Kuffer, M. *et al.* IDEAMAPS: Modelling Sub-Domains of Deprivation with EO and AI. in *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium* 1562–1566 (2024). doi:10.1109/IGARSS53475.2024.10642494.

45. Cámara-Menoyo, C., Porto de Albuquerque, J., Suchomska, J., Tregonning, G. & McInerny, G. Co-designing grounded visualisations of the Food-Water-Energy nexus to enable urban sustainability transformations. *Environ. Sci. Policy* **154**, 103712 (2024).

46. Hafner, S. *et al.* Towards an Indicator-Based Morphological Informality Model for Sub-Saharan Africa Using Open Building Footprint and Road Data (Version 1). *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **XLVIII-M-7–2025**, 221–228 (2025).

47. Sirko, W. *et al.* Continental-Scale Building Detection from High Resolution Satellite Imagery. Preprint at https://doi.org/10.48550/arXiv.2107.12283 (2021).

48. Wang, J. *et al.* EO + Morphometrics: Understanding cities through urban morphology at large scale. *Landsc. Urban Plan.* **233**, 104691 (2023).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.docx