

Communication

# Earth Observations and Statistics: Unlocking Sociodemographic Knowledge through the Power of Satellite Images

Paloma Merodio Gómez <sup>1</sup>, Olivia Jimena Juarez Carrillo <sup>1,\*</sup>, Monika Kuffer <sup>2</sup>, Dana R. Thomson <sup>3</sup>, Jose Luis Olarte Quiroz <sup>1</sup>, Elio Villaseñor García <sup>1</sup>, Sabine Vanhuyse <sup>4</sup>, Ángela Abascal <sup>5</sup>, Isaac Oluoch <sup>6</sup>, Michael Nagenborg <sup>6</sup>, Claudio Persello <sup>2</sup> and Patricia Lustosa Brito <sup>7</sup>

- <sup>1</sup> Instituto Nacional de Estadística y Geografía (INEGI), Aguascalientes 20276, Mexico; paloma.merodio@inegi.org.mx (P.M.G.); jlolarateq@gmail.com (J.L.O.Q.); elio.villasenor@inegi.org.mx (E.V.G.)
  - <sup>2</sup> Geo-Information Science and Earth Observation (ITC), University of Twente, 7514 AE Enschede, The Netherlands; m.kuffer@utwente.nl (M.K.); c.persello@utwente.nl (C.P.)
  - <sup>3</sup> Department of Geography and Environmental Science, University of Southampton, Southampton SO17 1BJ, UK; drt1g15@soton.ac.uk
  - <sup>4</sup> Department of Geosciences, Environment and Society, Université Libre de Bruxelles (ULB), 1050 Brussels, Belgium; sabine.vanhuyse@ulb.be
  - <sup>5</sup> School of Architecture, Navarra Center for International Development, University of Navarra, 31009 Navarra, Spain; maabascal@unav.es
  - <sup>6</sup> Department of Philosophy, University of Twente, 7522 NB Enschede, The Netherlands; i.o.oluoch@utwente.nl (I.O.); m.h.nagenborg@utwente.nl (M.N.)
  - <sup>7</sup> Politechnic School of Federal, University of Bahia (UFBA), Bahia 40210-630, Brazil; patricia.brito@ufba.br
- \* Correspondence: jimena.juarez@inegi.org.mx



**Citation:** Merodio Gómez, P.; Juarez Carrillo, O.J.; Kuffer, M.; Thomson, D.R.; Olarte Quiroz, J.L.; Villaseñor García, E.; Vanhuyse, S.; Abascal, Á.; Oluoch, I.; Nagenborg, M.; et al. Earth Observations and Statistics: Unlocking Sociodemographic Knowledge through the Power of Satellite Images. *Sustainability* **2021**, *13*, 12640. <https://doi.org/10.3390/su132212640>

Academic Editor: Anna Visvizi

Received: 23 August 2021

Accepted: 16 September 2021

Published: 16 November 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** The continuous urbanisation in most Low-to-Middle-Income-Country (LMIC) cities is accompanied by rapid socio-economic changes in urban and peri-urban areas. Urban transformation processes, such as gentrification as well as the increase in poor urban neighbourhoods (e.g., slums) produce new urban patterns. The intersection of very rapid socio-economic and demographic dynamics are often insufficiently understood, and relevant data for understanding them are commonly unavailable, dated, or too coarse (resolution). Traditional survey-based methods (e.g., census) are carried out at low temporal granularity and do not allow for frequent updates of large urban areas. Researchers and policymakers typically work with very dated data, which do not reflect on-the-ground realities and data aggregation hide socio-economic disparities. Therefore, the potential of Earth Observations (EO) needs to be unlocked. EO data have the ability to provide information at detailed spatial and temporal scales so as to support monitoring transformations. In this paper, we showcase how recent innovations in EO and Artificial Intelligence (AI) can provide relevant, rapid information about socio-economic conditions, and in particular on poor urban neighbourhoods, when large scale and/or multi-temporal data are required, e.g., to support Sustainable Development Goals (SDG) monitoring. We provide solutions to key challenges, including the provision of multi-scale data, the reduction in data costs, and the mapping of socio-economic conditions. These innovations fill data gaps for the production of statistical information, addressing the problems of access to field-based data under COVID-19.

**Keywords:** data cubes; deprivation; urban poverty; slums; data ecosystem; statistics

## 1. Introduction

The variations of urban socio-economic conditions, in particular, the proliferation and expansion of slums, are a global concern [1]. The world's slum population is estimated to reach 3 billion people by 2030 [2]. However, these statistics come with many inherent uncertainties, for instance, due to large data gaps, dated statistics, or difficulties accessing

areas for field verification [3]. Hence, the systematic quantification of socio-economic conditions (and slums) requires methods to capture their spatial patterns in a consistent manner to support pro-poor policies, upgrading programmes and SDG 11 monitoring. Therefore, an increasing number of research activities aim to develop robust tools and techniques for collecting information on socio-economic conditions, and to fill gaps in publicly available datasets [4].

It is important to clarify that slums and informal settlements are not synonymous. In many settings, slums refer to urban areas without essential services and adequate living conditions. Informal settlements are those settlements developed outside legal regulations. Typically, slums are also informal settlements, but this may not always be the case; it is possible to find informal settlements with wealthier conditions, and slum conditions are observable by a field survey, while informality is corroborated by administrative records.

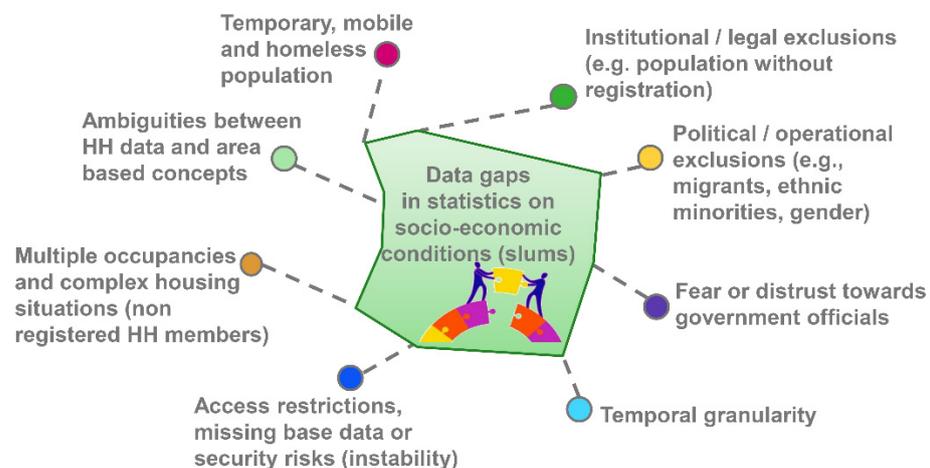
Although slum conditions are visible, their exact locations are often missing in official maps; additionally, local slum definitions might differ, and thus, boundaries are often fuzzy. In general, it is necessary to consider multiple dimensions to characterise slum conditions. The approach that is commonly used to monitor SDG 11 relies on aggregated household assets data from censuses and surveys [5], and produces imprecise estimates of people living in slums, informal settlements, and other deprived urban areas. Census data are routinely collected in many countries, nevertheless, they are costly and require huge logistics to visit each and every single household in the country, making it difficult to routinely acquire data (commonly every 10 years). The available data are aggregated to relatively small areas to protect individual privacy. However, even if the information is available every 10 years, the rapid socio-economic and physical changes on the ground, especially in slums, make data often outdated at the moment they are published. Furthermore, the degree of data aggregation can also introduce a bias, also referred to as the modifiable areal unit problem [6]. Other limitations of this approach are that poor households do not necessarily congregate in deprived areas, and household poverty (e.g., unimproved structure, crowding, or inadequate water and sanitation) represent different phenomena than area-level deprivation (e.g., lack of infrastructure, pollution, social exclusion, or limited access to basic services) [7,8].

Furthermore, the recent COVID-19 pandemic has caused additional challenges for collecting information on the ground, urgently increasing the use of alternative data sources to capture socio-economic characteristics. National Statistical Organizations (NSOs) are being challenged by serious disruptions to their usual work. A survey launched in May 2020 (for details, see Appendix A) shows the heavy impact of COVID-19 on NSOs (work carried out by the aegis of the United Nations Statistical Division (UNSD), the World Bank's Development Data Group and the UN Regional Commissions) [9]. The questionnaire was sent to 218 NSOs, from which 122 responses were received. The survey revealed that 65% of NSO offices are fully or partially closed. Face-to-face data collection has been impacted 96% of responding offices, with 69% stopping it completely [9]. Fieldwork had to be postponed or stopped. Meanwhile, for all censuses that were planned to be carried out during 2020, preparatory activities were impacted in 55% of cases. For example, the Brazilian census (planned in 2020) has not been conducted, which will dramatically impact SDG monitoring.

A key challenge in urban areas is the identification of disadvantaged population locations to strengthen social programs and to promote social equity [10]. However, even before COVID-19, the corresponding data were often not available or were outdated. For instance, traditional poverty measures (e.g., based on household surveys) are available at low frequency (at best every 5 to 10+ years). Due to their sampling scheme, they are only representative of coarse spatial granularity. Additionally, for many areas, they are not available; for example, between 2002 and 2011, 57 countries out of 155 (countries covered by the World Bank poverty data), had less than two data points, which is the minimum required for properly measuring poverty during a ten-year period [10]. Furthermore, data are often aggregated at geographically arbitrary units that average and hide poor areas.

Data innovations are key in support of evidence-based decision-making and to monitor the implementation of policies [10–12].

On a global scale, the causes of data gaps about slums are summarised in Figure 1. They relate to the difficulties to include temporary and mobile population (e.g., migrant workers), institutional or legal exclusion of populations without formal registration, political or operational exclusions (e.g., because of unsolved migrant status or the systematic exclusion of women or ethnic minorities in some countries), distrust of surveyed populations towards government officials (e.g., fear of eviction), low temporal granularity of censuses, limited access by data collectors (e.g., due to insecurity or simply not having records of new settlements in remote peri-urban areas), complex housing conditions that are not fully recorded (e.g., multiple occupancies or sub-renting) and general differences between the household (HH) level slum definition and an area-based understanding of deprivation [13]. For many of these issues data gaps, EO could be of great utility, offering spatially detailed data with high temporal granularity.



**Figure 1.** Summary of causes of data gaps in official statistics on socio-economic conditions and slums in particular (summarised from [14–16]).

The EO industry is undeniably exploding because EO presents an opportunity to build consistent time series that can fill data gaps. Furthermore, the availability of free imagery from satellite constellations like Landsat, Sentinel, or MODIS has democratised access to global and timely satellite imagery. Meanwhile, cloud computing providers like Amazon Web Services (AWS) and Google Cloud store satellite data for free, which is further accelerating the global usage of these EO data.

Diverse applications of EO data have been linked to various dimensions of global agendas. For example, [17] listed cases related to urban development (land use, housing, transportation, water, air quality, energy, climate change and others) that can be monitored using EO and that are relevant to different global frameworks (Sustainable Development Goals (SDGs), New Urban Agenda (NUA), Sendai Framework and Paris Agreement). Users can access daily high-resolution satellite images of the entire Earth, allowing the rapid production of land cover and land use information.

High- and very-high-resolution images allow the automated extraction of specific objects (e.g., buildings, cars, waste piles). Such object-level information can be used to approximate socio-economic conditions of an area (e.g., absence of cars as a poverty indicator). Additional information, for example, the presence of night-time lights (captured by night-time images), serve as an indicator of economic activities. However, these alternative methods to produce socio-economic and demographic estimates in data-scarce environments present several challenges related to the spatial coverage (e.g., limited access to high-cost very-high-resolution (VHR) imagery), resolution (e.g., the coarse resolution of night-time light imagery), comparability (e.g., combining different sensors for multi-temporal analysis)

and rapid dynamics that require high temporal granularity in monitoring. Thus, neither is the potential of EO to fill data gaps in official statistical databases sufficiently explored, nor are the most suitable EO methods and data sources in terms of cost–benefit definition [18].

The recent broad availability of timely satellite images unveils yet a different challenge: there are large amounts of data to be analysed, for which traditional techniques are not enough. In 2012, the “deep-learning revolution” opened up an entirely novel frontier. Modern Machine-Learning boosted techniques such as the detection and automatic counting of objects, semantic segmentation, and generic image classification [19]. For example, deep learning has proven to be incredibly versatile [20]. An essential role was played by large benchmark datasets for image analysis, such as the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), which has served as a testbed for the development of deep learning networks in computer vision.

However, deep-learning architectures as such are not sufficiently operational to fill data gaps for NSOs. For example, they are typically set to small images with three VIS “channels” (red, green, and blue), stored as image files (e.g., PNG, JPEG) [19]. Thus, these architectures are not set to the use of satellite images, which are often large files, including a geo-reference, with many multispectral channels [20]. Recent research efforts are aimed at designing deep network architectures tailored to specific EO data types. For instance [21], Bergado et al. introduced a multiresolution fully convolutional network developed for the analysis of VHR images [18]. The network (dubbed FuseNet/ReuseNet) is able to fuse the panchromatic and multispectral images and perform land-cover classification considering contextual label information in one end-to-end learning framework [21]. Other networks have been specifically designed for the analysis of Synthetic Aperture Radar (SAR) data [22,23]. While the former CEO of Google, Eric Schmidt, stated that image recognition is a “solved problem” [19], this is not true for the EO deep-learning community, with many unsolved challenges for even basic workflows. For example, large archives of (multi-temporal) ground-truth data are missing for deep-learning algorithms, which generally are very training intensive. Unlike regular image recognition, the method to create a ground-truth dataset (whether it is collecting field data or carried out with specialised work of a photo-interpreter) is a decision with all sorts of consequences, such as the quality of the results or the cost of the study itself.

During a conference on EO and Statistics organised by the National Institute of Statistics and Geography of Mexico and the University of Twente (recording available <https://slummap.net/index.php/2020/08/07/webinar-earth-observations-statistics/> (accessed on 15 January 2021)), it was showcased that EO-based proxies of socio-economic and demographic data can provide timely and relevant information, including extensive spatial coverage as well as multi-temporal data. The aim of this paper is to showcase new avenues to unlock the full potential of EO in support of generating sociodemographic knowledge, focusing in particular on the provision of data on slums. This challenge exhibits three major dimensions where innovation is urgently required: (1) defining a suitable aggregation level of EO data in terms of multi-scale data provision; (2) a critical analysis of data costs/benefits; and (3) the innovation in methods to unlock the full potential of EO. Section 2 of this paper provides an analytical framework that highlights the potential of EO data. We then showcase in Section 3, using several cases, how these challenges can be addressed by the use of EO. Next, we discuss the benefits and limitations of the approach also linked to data needs during COVID-19 and finally concluded in Section 5.

## 2. Materials and Methods

As mentioned, an alternative to census data is the application of machine learning and, in particular, deep learning using EO data. Using several cases, we will highlight advancements of EO to fill existing data gaps for NSO.

The first step in any machine learning application is to construct a training dataset. That is a set of examples of the objects we want the algorithm to make predictions about. These objects have to be labelled, which means that we know a priori the value of the

characteristic that we want the algorithm to predict. No matter the machine learning method used, representative and vast training datasets are required to obtain accurate results, particularly when predicting complex classes (e.g., socio-economic conditions), which are not as evident as the presence of vegetation, for example.

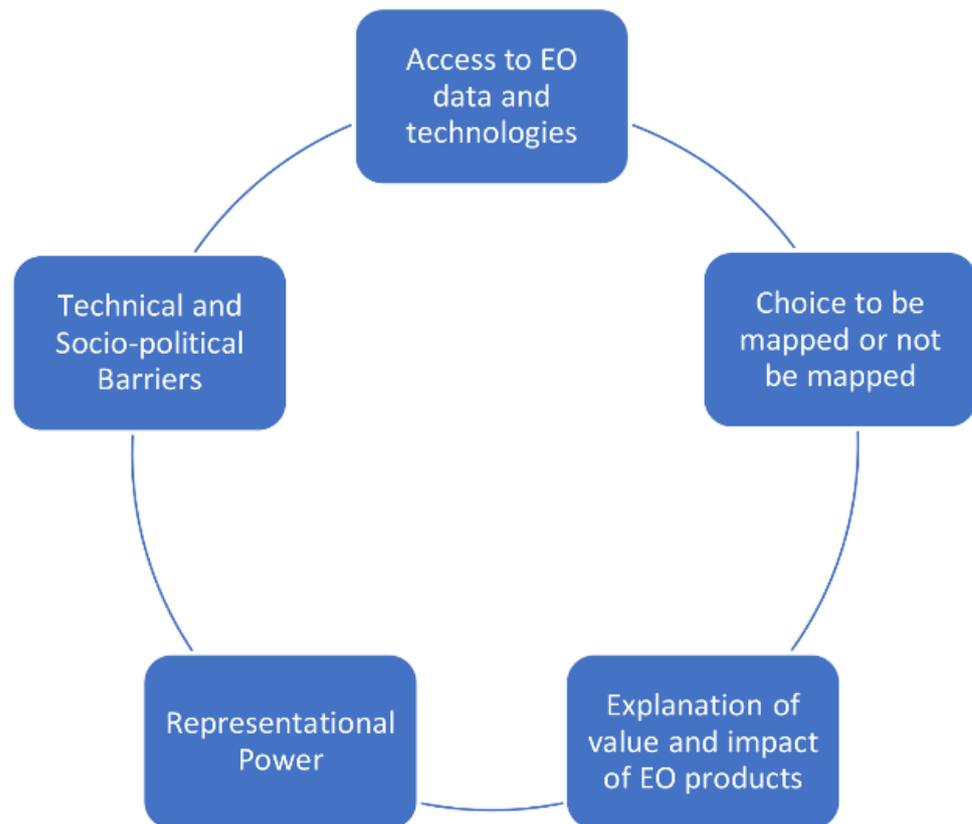
The use of deep learning algorithms to predict not only the location of slums but also the slum severity (or deprivation level) with EO data (which is an open problem) will be highlighted by selected cases. There is relevant progress in this particular use of EO, e.g., [24] showed that deep learning and transfer learning techniques allow for processing VHR images, reaching an overall accuracy of 94.2; nevertheless, very few studies analyse physical differences between slums [25,26]; most studies assume that slums are homogenous (e.g., [27,28]). One exciting and challenging approach to improve model results is to combine distinct types of images [29]. Additionally, the analysis of dynamics of slums is still challenging, hindered by limited access to commercial VHR images (e.g., around \$20 per km<sup>2</sup>) [30]. It is also worth noting that even when data and computational challenges are addressed, it is not clear to most EO experts how data on slums can be made available without putting the privacy and security of slum residents at risk.

### 2.1. Ethics for Urban Poverty Mapping

The case studies (Section 3) show how privacy can be optimally addressed. When implementing machine learning methodologies and selecting a framework to use EO as input for mapping slums, “Mapping Ethics” should entail an assessment of five components: access, choice, value, power, and barriers (Figure 2). Firstly, the level of access to EO data and technologies affects the capacity of those mapping and those being mapped to interact with, challenge or agree on the representations made by EO technologies. This capacity depends on whether the activity of mapping is inclusive or exclusive. However, this capacity also depends on whether those being mapped have a choice in being mapped or not, as well as the technical and socio-political barriers that may exist. The choice to be mapped reflects the existence or lack of a dialogue between those in urban deprived areas, researchers, municipal authorities and international NGOs. This dialogue should increase awareness among those being mapped about the value and impacts of the data used to design EO products, a dialogue that transitions those being mapped from data subjects to potential data citizens [31]. However, this dialogue should also feature the choice of those being mapped on how they would like to be mapped. For instance, the youth in Brazil, “have developed an aversion to any location-disclosure behaviour, including tagging, to protect their personal safety” [32] due to gang presence in favelas. There is, therefore, a need to pay attention to international as well as local data disclosure and privacy protection concerns [33,34] which influences what kinds of visualisations and platforms are used to represent the information on those living in urban deprived areas.

More so, this dialogue on how and whether to be mapped can be impeded by technical and socio-political barriers and the representational power of EO products. In its technical use, representational power points to the ability of computational neural networks to assign labels to specific instances of defined classes (in this case, boundaries of “slum”/“non-slum”). However, the socio-political dimension of representational power, relates to the impact of these designations on the lives of those being mapped. EO technologies and products do not operate in a scientific vacuum; they can affect how well the communities of those being mapped are seen at various levels (e.g., by researchers, municipal authorities and international organisations); and the decisions made at these levels can lead to improving not only the technical representation in terms of accuracy [15,25,35–38] but also socio-political representation such as recognition of rights [39,40]. However, the technical, as well as socio-political dimensions of representational power, can be impeded by a number of barriers, (i) time taken to set up EO technologies and platforms, (ii) cost of acquiring EO data and technologies, (iii) skills to understand and manipulate EO data, (iv) interest from State authorities in using EO data and products [41], and (v) trust between those being mapped and those doing the mapping ([34], p. 3). These components frame

the potential as well as challenges in setting up a “Mapping Ethics” in the context of urban poverty mapping. In light of data privacy and Mapping Ethics, the development of a data ecosystem needs a careful design which considers: the protection of vulnerable communities from being further stigmatised; the dialogue with communities in terms of awareness and access to data; and local understanding of which data should not be shared in such a system.

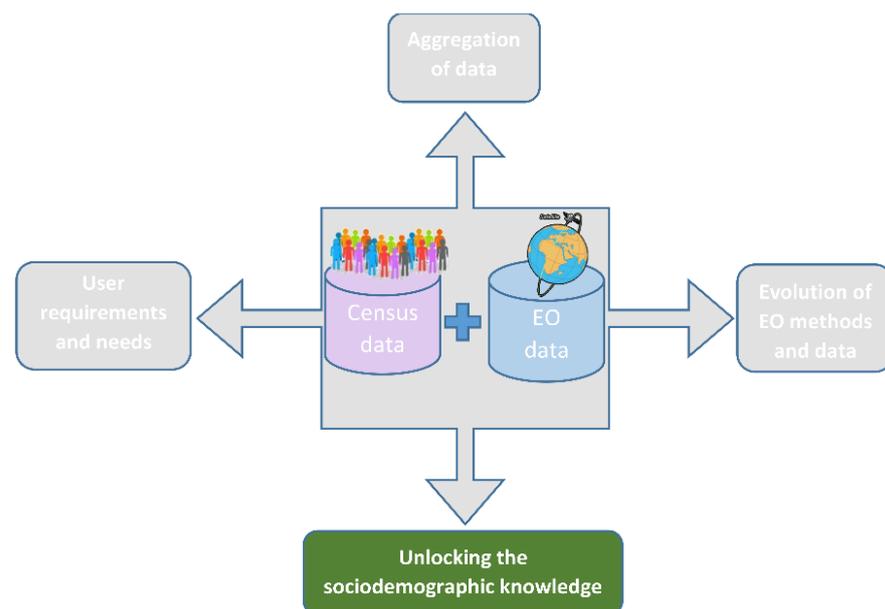


**Figure 2.** Mapping ethics and privacies in the context of slum mapping.

## 2.2. Case Study Selection

In the following section, three cases are used to show how major challenges can be solved which would allow unlocking the potential of EO data (Figure 3). First, to solve data aggregation, one solution is the use of regular area grids. The use of a gridded mapping system allows data producers to provide consistent data with improved spatial resolution and near-continuous coverage (Section 3.1). Grid-based maps show a continuous distribution across a defined (usually rectangular) area, providing an efficient solution to the problem caused by geographically arbitrary units such as administrative areas, in which averages may hide poor areas. Grids can be up-scaled to larger geographic units such as administrative boundaries. Hence, grids offer a consistent and standardised approach to spatial data collection, analysis or visualisation. Furthermore, considering the growing availability of machine learning for the analysis of satellite images (e.g., for more timely information), the discrete nature of grid-based datasets opens up new possibilities. The urgency of providing consistent base data to support rapid yet informed decision-making during challenging situations, such as COVID-19, has been stressed by [42]. The development and role of gridded mapping systems are highlighted in two cases (one national and one continental) to highlight development stages and reflect on data aggregation and mapping ethics. Second, most studies on mapping slums work with VHR commercial images, leading to very high data costs (Section 3.2). However, there is an insufficient critical analysis on the cost–benefit relation between information needs and spatial resolution (towards low-cost mapping systems). Therefore, the second case

illustrates the role of low-cost data for the basic mapping of slums versus more detailed slum area characterisation. The case emphasises methodological innovations required to address the challenges in selecting appropriate EO data, as well as the development of suitable methods to inform end-users and decision-making supporting the creation of public policies. This case focuses on one city (Nairobi, Kenya) to show the potential of different EO data at different scales of analysis to support local data needs. The third case (Section 3.3) looks into the complexity of measuring the intra-urban variability of socio-economic conditions, departing from the mapping focus on slums by understanding urban multi-deprivation poverty as a continuous phenomenon and asking for mapping approaches that do not point to slums as stigmatised areas. This case uses a rich dataset that combines EO data with socio-economic data to model deprivation at the city scale for the case of Bangalore, India. The combination of cases will provide the way forward towards unlocking the sociodemographic knowledge with EO data and addressing major urban data gaps.



**Figure 3.** Framework towards unlocking the sociodemographic knowledge with EO data.

Therefore, the following section highlights innovations towards solving these three challenges ((i) data aggregation (gridded mapping), (ii) development of low-cost mapping systems, and (iii) intra-urban variability of socio-economic conditions. Furthermore, the cases also deal with privacy and data ethics. These challenges were identified based on a scoping literature review carried out for the period 2017–2020 (including some preprints/versions that were published in 2021). The cases show how innovations in EO can contribute to policy-relevant information, with the aim to provide stakeholders (ranging from the local to national levels) with the urgently required information for evidence-based policymaking.

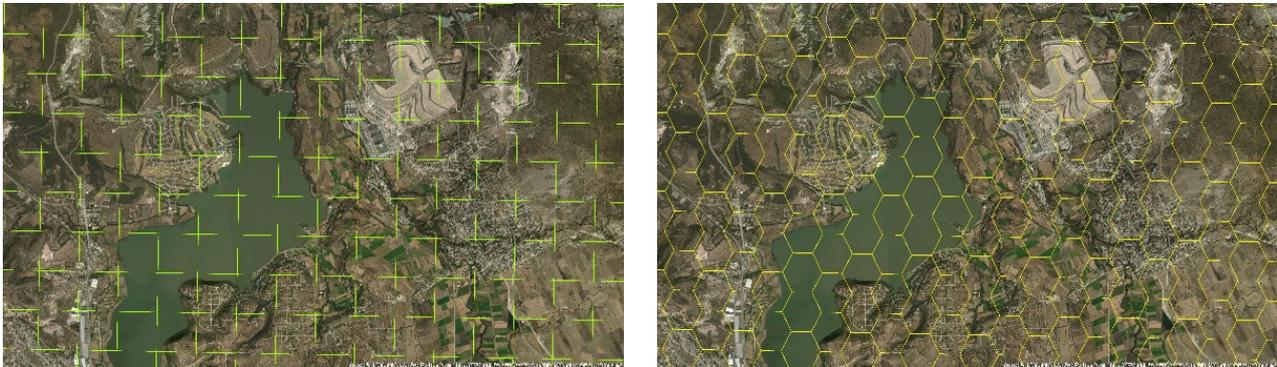
### 3. Cases Studies: Unlocking the Sociodemographic Knowledge with EO-Methods

The selected cases show how advancements in EO can fill data gaps for NSOs and support SDG 11 monitoring.

#### 3.1. Gridded Systems for Collecting Sociodemographic Data and the Role of Data Cubes

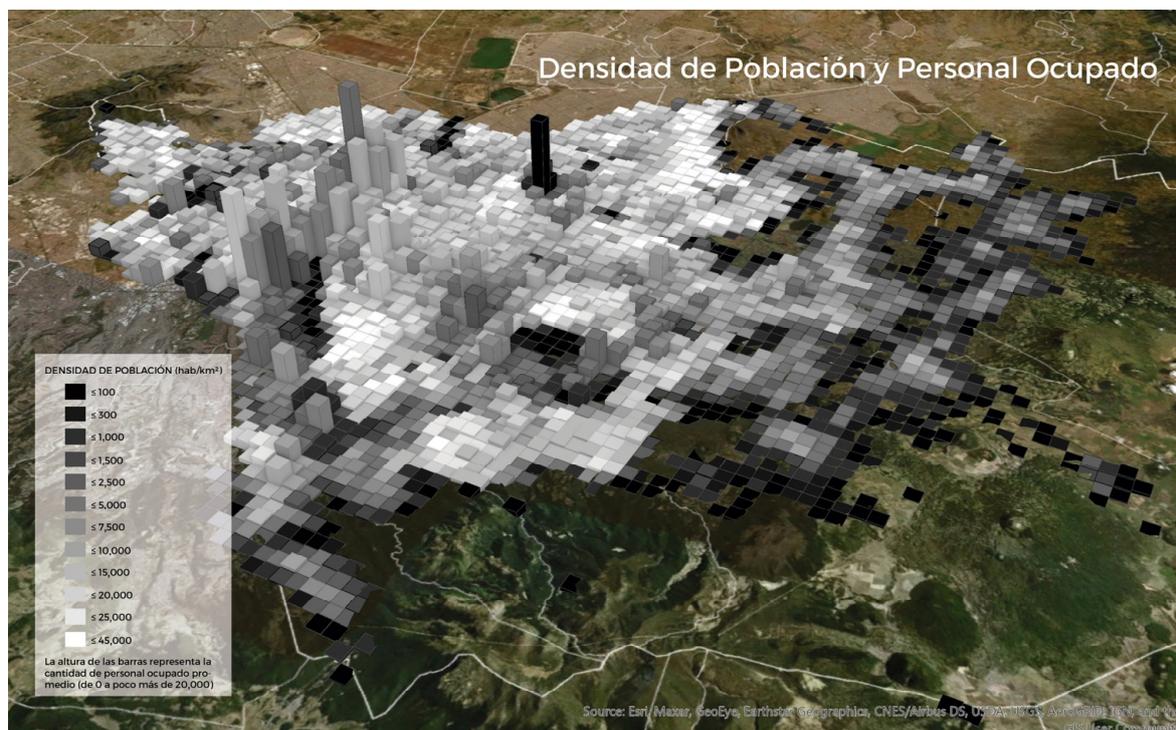
To link EO data with the existing statistical data, data cubes as well as data ecosystems, there are innovative solutions. In the context of data sciences for EO, and for lessening the difficulty of comparing, merging and sharing data that comes from sources with different geographies, the use of grids [43] has emerged as a good option. A grid is a

set of continuous geo-spatial units—cells—which have the same geometry as well as a homogeneous distribution of a thematic content (population, housing, vegetation, slope of the terrain, etcetera) (Figure 4).



**Figure 4.** Two examples of gridded data, regular squared grid (left) and hexagonal grid (right).

The purpose of Mexico’s grid is to provide accurate training data to their Data Cube as it provides a standard unit throughout time (instead of the varied statistical units that are traditionally used to report the data), in order to take advantage of its descriptive and predictive capacities (Figure 5). Thus, Mexico’s grid was designed under geometric and cartographic specifications of the Mexican Geospatial Data Cube. This allows summarising the socio-demography, the economic and geographical data of the country at the grid cell level. For a particular grid, it was necessary to proportionately distribute 2.3 million city blocks to the grid cells, consisting of more than 112 million people, a little more than 28 million inhabited private households and 4.3 million economic establishments (INEGI 2010), as well as several geographic layers, to solve the urban–rural dichotomy through several density indicators as well as service availability inside the households.



**Figure 5.** Gridded map of the population density and employment of Mexico City, 2015 (Source: INEGI).

Similarly, the Integrated Deprived Area Mapping System (IDEAMAPS) is also currently implementing a gridded mapping approach in the form of a data ecosystem that combines data to understand urban deprivation (the first pilot of the system is available <https://ideamapsnetwork.org/> (accessed on 20 August 2021). IDEAMAPS was conceived in 2019 to produce routine, accurate maps of urban deprivation in Low-and-Middle-Income Countries (LMICs) by integrating the strengths of existing, silo-ed “slum” mapping approaches [11]. The IDEAMAPS Network officially launched in 2020 with funding from a UK Research and Innovation grant [11].

The use of EO and other spatial data in machine learning models is well suited to map area-level phenomena; however, models need to be combined with community-based data to be validated and to be locally relevant [18]. Therefore, the IDEAMAPS approach combines citizen-generated, EO data, census, survey, social media, web-scraped, and other data to produce a common, dynamic, accurate map of deprived urban areas in support of evidence-based policy-making and local planning and upgrading so that all cities can become equitable, healthy, and prosperous. The underlying IDEAMAPS data ecosystem facilitates fair exchanges of data among stakeholders and evolves improved accuracy over time (Figure 6). While a plethora of open datasets exists, they are dispersed, aggregated, inconsistent, or require GIS skills to access and preprocess (e.g., to clean data inconsistencies), and IDEAMAPS aims to remove these barriers. Neighbourhood-specific data are vital to community leaders for advocacy, gaining legal tenure, development, and crisis response. In IDEAMAPS, government stakeholders validate models in exchange for neighbourhood-level data summaries and deprivation maps classified into slum/non-slum areas (or multiple levels of deprivation) for SDG 11 monitoring, urban planning, budget planning, and other decision-making. When government officials participate in the modelling process, they gain understanding and the ability to certify outputs as official data, a current barrier to government use of open data.

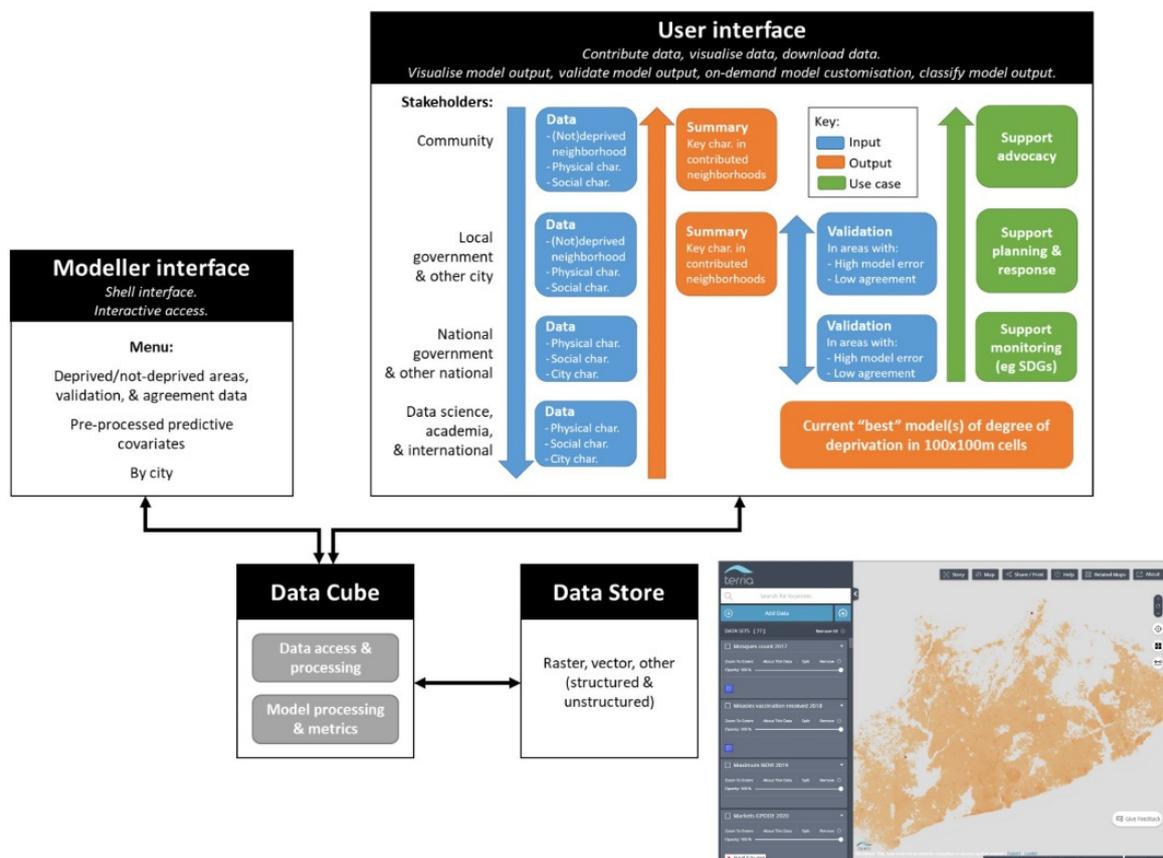


Figure 6. The IDEAMAPS data ecosystem and pilot gridded mapping system for Accra, Ghana.

By meeting community, government, and other key stakeholder needs, IDEAMAPS fosters a living data ecosystem of continually contributed training and covariate data defined by local context experts, and the ability to incorporate an important parameter into deprivation models: level of agreement among local experts about what is, and is not, a deprived area in a given city. To protect vulnerable communities from eviction, harassment, fines, and other negative consequences, all IDEAMAPS data visualisations are aggregated to a 100 m × 100 m grid system to obfuscate the exact boundaries of vulnerable communities. This also serves to anonymise sensitive information and streamline data to match existing urban datasets (e.g., GHSL, WorldPop) for data scientists who contribute modelling innovations to the IDEAMAPS ecosystem.

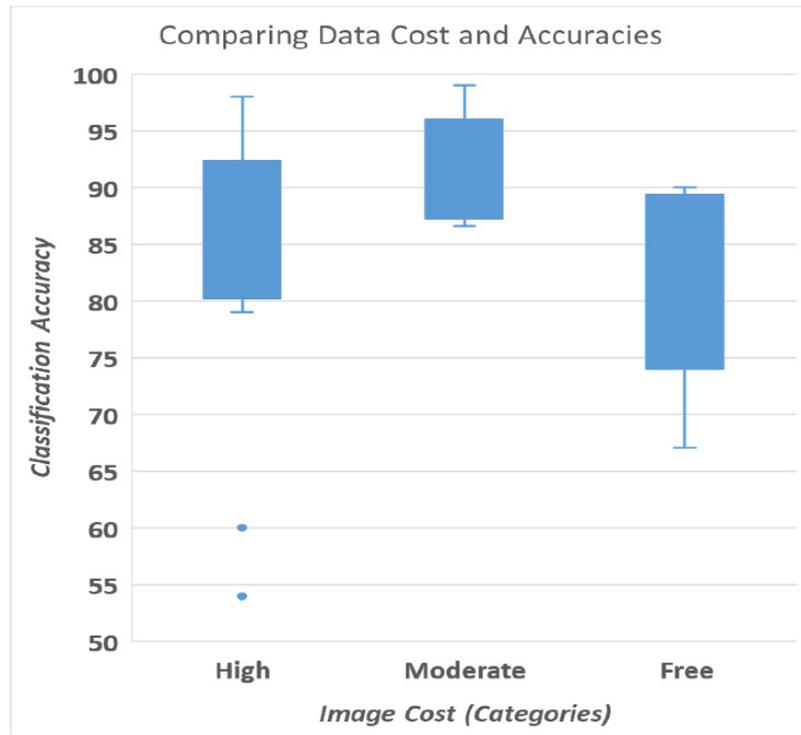
Both examples highlight the possibilities of the grid as an artefact for providing training data for taking advantage of EO via its predictive capacity in a data cube, as well as its importance as a useful technique for viewing and sharing the results. As such, it focuses on improving the interpretation, comparability, merging and exchanging of geo-spatial data that comes from different sources. Gridded systems protect vulnerable groups by not showing exact boundaries while still providing data access that is useful for interaction and cooperation with diverse citizen groups.

### 3.2. The Role of Low-Cost Data for Mapping Slums

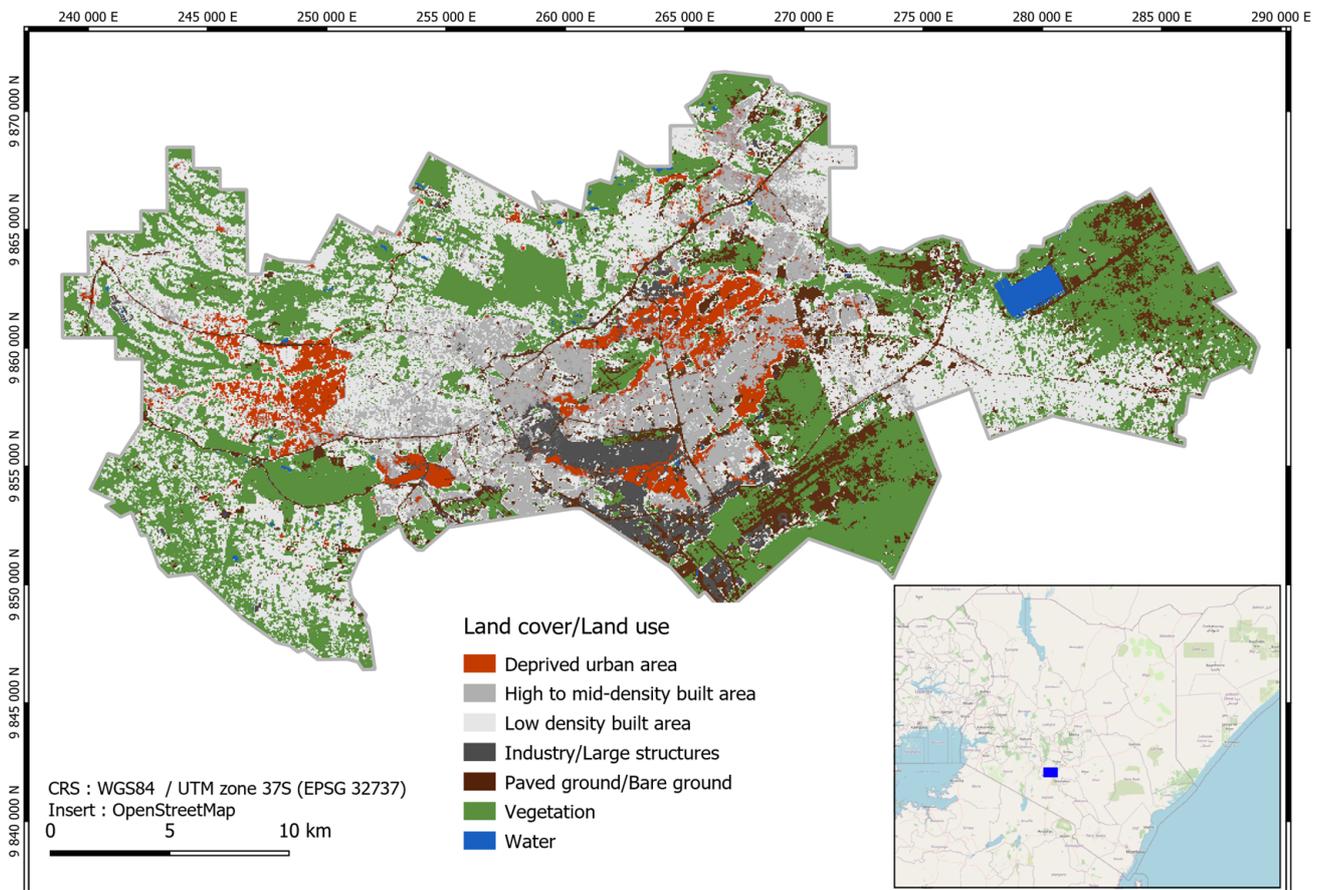
The majority of studies on mapping deprived urban areas (e.g., slums) work with commercial VHR imagery. Analysing all studies from recent years found in Scopus (N = 28) [15,24,25,28–30,36,41,44–63], more than 50% work with VHR optical imagery, followed by 30% high resolution (HR) optical (e.g., Sentinel-2) or Google Earth imagery and bit less than 20% with SAR imagery (both VHR and HR). The dominance of VHR imagery is also contributing to the dilemma of most studies developing and testing methods on very small areas (usually few km<sup>2</sup>), which is insufficient to reflect on the potential of integrating EO and statistical data, both typically available for large area coverage.

In general, within the EO community, there is an understanding that VHR imagery is necessary to reveal complex urban patterns and socio-economic information [21]. To understand the relation between image cost and resolution, a literature review was carried out on slum mapping studies published (including preprints) between 2017–2020. In the retrieved literature (N = 28), the relationship between accuracies and EO data costs was analysed (Figure 7). The image costs were grouped into three classes (where “High” means more than 10 EUR/km<sup>2</sup>, “Moderate” meaning 10 EUR/km<sup>2</sup> or less, and “Free” imagery without costs), there is no strong relation between high accuracies and high imagery cost. The variations within the three classes (of image cost) can be explained by the use of different machine learning algorithms and the different complexity of the cities. In general, free imagery (mainly Sentinel and Google Earth) achieved a bit lower accuracy, but reached an accuracy of 80% on average. Thus, they are a promising alternative to avoid image data costs. As an example, in the SLUMAP project (<https://slumap.ulb.be/>, accessed on 10 June 2021), gridded maps were produced using Sentinel-1/2 (Figure 8).

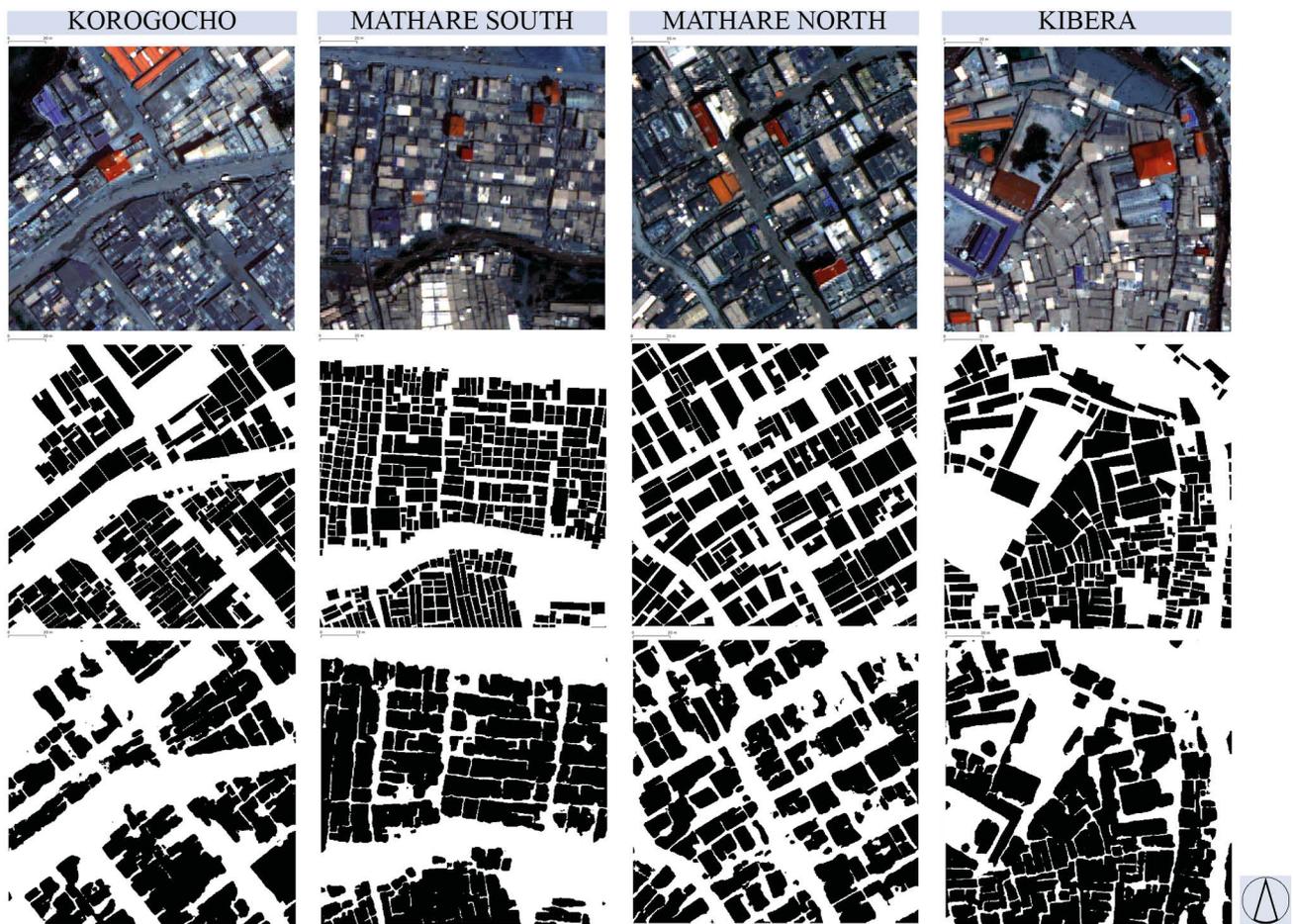
Several studies (e.g., [63–65]) have confirmed that VHR resolution images (of 1 m and below spatial resolution) are not optimal for mapping deprived areas at city scale as they are impacted by too-high-level details, which is adding noise to the classification. Thus, an optimal resolution for city-scale maps depends on the urban morphology and typically vary around 2–5 m to capture deprived areas. However, this conclusion differs when it comes to mapping objects (e.g., buildings). For this purpose, VHR imagery is fundamental to map building objects that are required for many urban morphological analysis purposes. A resolution definitely below 0.5 m is required to map buildings in deprived areas. However, in very densely built-up urban areas, such as deprived areas in Nairobi (shown in Figure 9), even 0.3 m is not sufficient to capture all roofs as separate objects (lower row Figure 8). Roofs in such areas are forming larger blobs which, even for visual interpreters, are difficult or even impossible to separate.



**Figure 7.** Comparing classification accuracies and image data costs (high > 10 EUR/km<sup>2</sup>, moderate ≤ 10 EUR/km<sup>2</sup> and free data) (number reviewed literature N = 28).



**Figure 8.** Gridded land-cover/land-use map (50 m × 50 m) of Nairobi highlighting deprived areas, produced from Sentinel-1 (SAR) and Sentinel-2 (optical) image features using a random forest classifier.



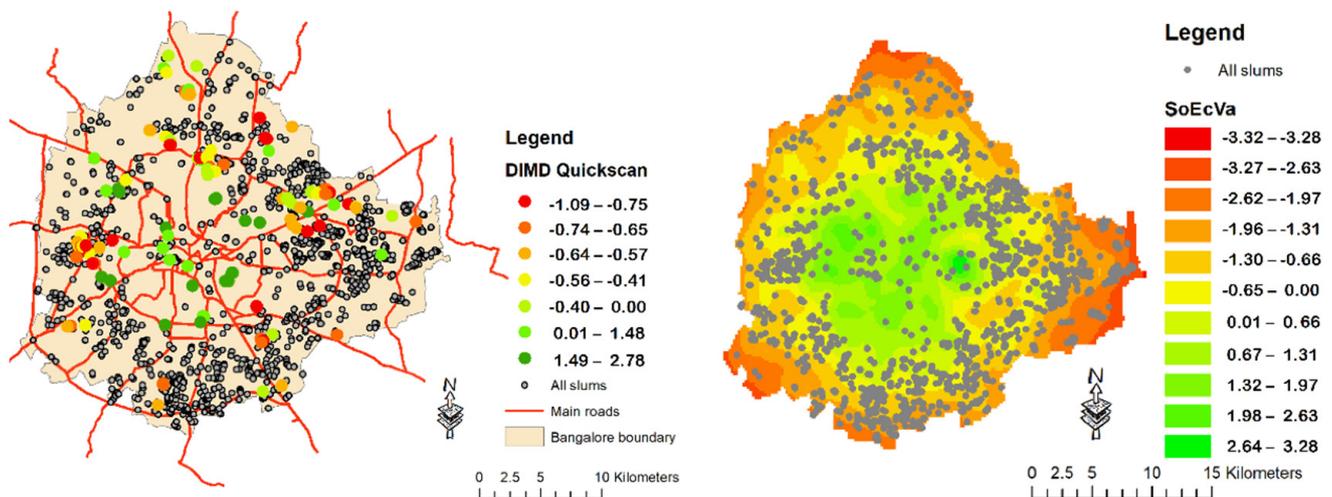
**Figure 9.** Building footprints in deprived areas in Nairobi using WorldView-3 and deep learning: WorldView-3 image (**upper row**), reference building footprints using visual interpretation (**centre row**), extracted buildings footprints using U-Net (**lower row**).

### 3.3. Socio-Economic Inequalities and Deep Learning

We stress the importance of adding area-level deprivation as compared to the household level concept of slums by UN-Habitat [13]. In recent years, many remote sensing studies [29,41,44,53,55,66,67] have highlighted the potential to map the physical aspect of deprivation, also referred to as morphological slums [68]. Such studies used mostly VHR imagery. There is a general agreement that machine learning methods have opened the door to include contextual information, besides spectral bands, to map deprived areas, e.g., by including textural features [27,69]. The introduction of advanced machine learning methods, i.e., deep learning techniques, such as convolutional neural networks (CNNs) [53] and fully convolutional networks (FCNs) [67], allowed for automatically learning of spatial, textural, and morphological features of deprived areas, which have been previously hand-crafted in more classical methods (e.g., random forest or support vector machine classifiers). However, most studies do not fully capture the complexity and variation of deprivation; they stay typically with a binary mapping output, which is resulting in high uncertainties along boundaries and for areas that are atypical slums [30,70], e.g., deprived areas of lower densities more regular outlines but still being deprived in terms of living conditions.

To deal with these problems, [25] introduced a data-driven approach to summarise multiple deprivation variables built on the asset vulnerability framework [71]. The socio-economic conditions at the city scale are modelled through a “data-driven index of multiple deprivations” (DIMD). A deep transfer learning approach based on a CNN is used to

capture the socio-economic variability. The deep network, pre-trained on a slum dataset, is tuned towards the prediction of the DIMD, addressing the associated regression problem. This framework, which allows assessing the degree of deprivation of surveyed deprived areas, is extended to model the variation of deprivation at the city scale (Figure 10 shows the example of Bangalore, India). The approach is extended using open-source data, i.e., OpenStreetMap (OSM), VIIRS night-time lights, WorldPop data, census, and data from the Demographic and Health Surveys (DHS). The results show that deep learning combined with open data allows not only to map the location, extent, and physical characteristics of deprived areas but also the variations of socio-economic conditions at the city scale. This allows departing from the problems of defining the boundaries between deprived and non-deprived areas and also opens a new perspective of including a more holistic understanding of deprivation at city, urban or regional scale. Such data, being built on remote sensing and open-source data, have the potential of being regularly updated, allowing a temporal analysis of socio-economic dynamics at an urban scale.



**Figure 10.** DIMD (data-driven index of multiple deprivations) for the QS locations (left) and socio-economic variability (SoEcVa) index (right).

Timely and spatially detailed information on socio-economic conditions providing city-wide information and their link to physical deprivation is an essential input for supporting urban planning and management, allowing timely interventions and impact analysis of policies. Furthermore, such information is essential for analysing vulnerabilities linked to natural hazards [72], climate change [73], serves as contextual information for health studies [13] and supports the monitoring of the progress and supporting the achievement of several SDGs, e.g., 1 and 11 [15].

#### 4. Discussion

In the reviewed literature (Section 1), mapping innovations that combine with big data, satellite images and machine learning are excelling [10]. However, their potential to complement official statistics (for NSOs) is largely unexplored. To solve the remaining challenges, the key is the promotion of collaboration and engagement between data scientists, NSOs and different potential user groups of EO-data and derived mapping products and hence, progress towards sustainability.

Summarising the major findings of the cases, and in general, of the EO and Statistics conference, we find these to be some of the important next steps to foster a continuous evolution in terms of data ecosystems and to achieve high-resolution sociodemographic maps and to advance poverty mapping and the monitoring of SDG 11 indicators:

- Strengthening local capacities to use and sustain these methods through methods that are easily reproducible and the promotion of training. In particular, those relevant

in case of crisis and disasters, providing readily available data for fast responses. For example, such data have been mostly absent for COVID-19 responses (further discussed in Section 4.1).

- Improving data infrastructure through data standards and formats that promote spatial interoperability, Analysis Ready Data (ARD), and scalable workflows like cloud computing [10]. For example, in support of local and national SDG monitoring (further discussed in Section 4.2).
- Accelerating validation of promising new approaches and assessing their cost/benefit and suitability for purpose, as well as account for uncertainties in data (further discussed in Section 4.3).

To better understand, monitor, promote and pursue this progress towards unlocking the sociodemographic knowledge with EO data, it is important to begin by performing a diagnosis. Each element will have a different impact depending on the particular region or institution considered but identifying these gaps and needs is an opportunity to create or reinforce the ideal collaborative partnerships that promote capacity building and education and, in this way, bridge the geo-spatial digital divide.

The main advances are focused on the innovation in data, ideas, techniques and methodologies. It is vital that these and other important developments, at the same time, align to those standards that allow them to spread worldwide, and yet still be able to be calibrated to the specificity of local needs.

Activities involved in the integration of statistical and geo-spatial information can evolve through the implementation of these and other techniques but will continue to address the need to deliver high-quality and timely decision-ready-data to the decision-makers; nevertheless, the better and faster they are informed, the quicker the right steps are taken in the road towards inclusive and sustainable development, and in particular towards achieving SDG 11.

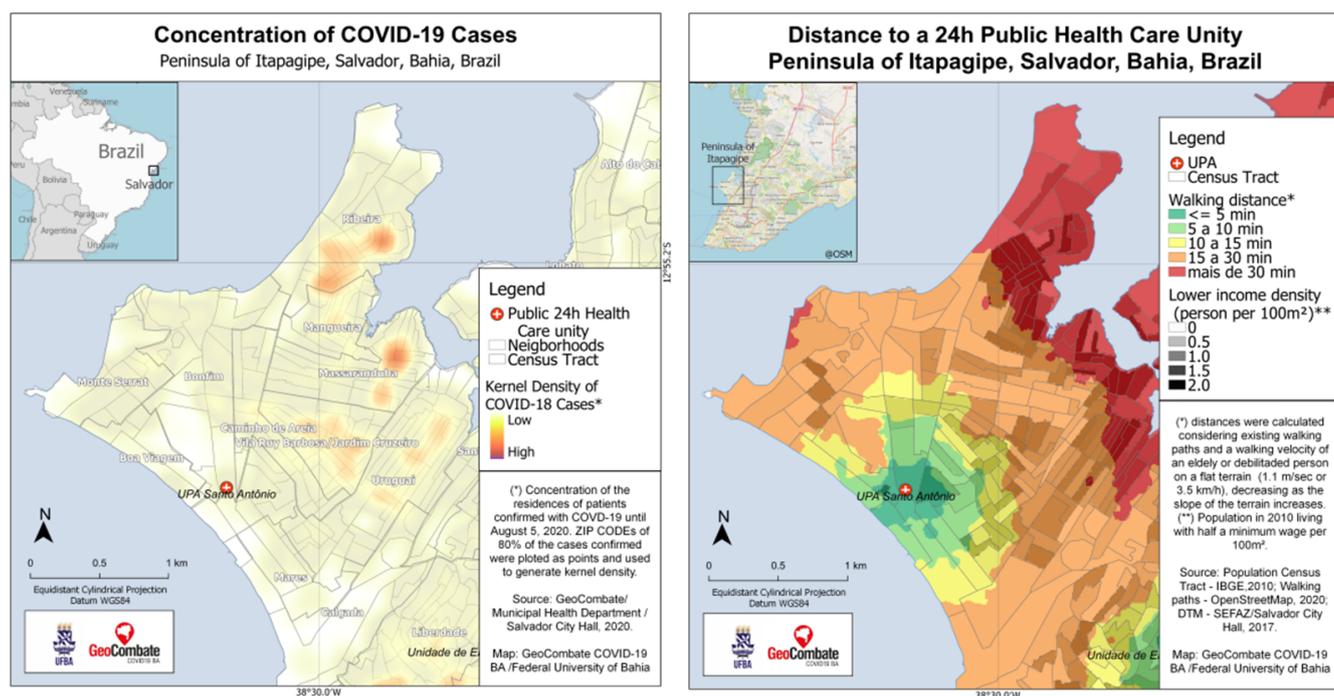
#### 4.1. EO Data for COVID-19 Responses in Slums

Geo-referenced socio-economic and demographic data are essential for any planning action, especially in emergency situations when that assistance must be fast and accurate. The availability of national data cubes or city-level data ecosystems would allow better management of such emergencies. With the declaration of the Sars-CoV-19 pandemic in March 2020 [74], there was an immediate and growing concern about its spread and impact in urban slums. Initially, the virus's arrival in LMICs through international travellers and first spread in wealthy neighbourhoods [75]. However, large inequalities that go from limited internet access, lack of tests and of medical care, to poor health conditions resulted in the fast spread beyond deprived communities (in particular slums) [76]. Mortality rates are showing to be high in deprived communities [77,78]. For example, in Salvador, Brazil, poor neighbourhoods had a mortality rate of 98.9 per 100,000 inhabitants, while in wealthier neighbourhoods, this rate was 86, considering the sum of cases from March to August 2020 [79].

Many observed challenges related to COVID-19 responses could be solved by the proposed EO-based data ecosystem. One first challenge relates to the location of risks, that is, the geo-referencing of positive cases in the intra urban space. The laboratory or medical records usually contain information on the patient's address, even though they may be incomplete, divergent or in an analogue format. For privacy reasons, the location of the residence should not be disclosed. However, few cities [42] publish their data in aggregation units such as streets, zip codes, grids up to 100 m or census tracts. The dissemination of information by neighbourhood, for example, requires the exclusion of non-residential areas from the analysis areas to avoid bias. For this, it is necessary to have updated data on urbanised areas and land use (ground information). Another problem is caused by the heterogeneity of large aggregation units. They generate distortions in the process of recognising the socio-economic profile, making epidemiological analyses difficult. Investigations end up being restricted to those who have access to the individual's

data, and results are not always disclosed to society. A second challenge is related to virus exposure, which is directly associated with social distance [80]. To support local strategies, estimates of household size (based on buildings footprints extracted from EO) and local land use classification are important to support dimension and distribution of awareness campaigns, testing, distribution of hygiene and protection items, disinfection, etc. The development of community-based reactions to the disease is the third challenge in slums. Once a person is suspected of having the disease, they need to be monitored, tested, isolated and treated and have every contact traced [81]. However, the problem may start at the moment of suspicion, if this person does not have and cannot pay for individual transport to the next health unit (most slums are far from such units—example in Figure 11. Therefore, primary health care, with a call centre or home visit, for example, becomes essential for monitoring the symptoms of COVID-19, but also for the control of pre-existing health conditions such as chronic diseases, e.g., diabetes, hypertension [82], other diseases that may be related to unhealthy and humid areas and homes, such as respiratory diseases [83] and infectious diseases (e.g., malaria or leptospirosis [84]).

Economic activities must go back to a “new normal” [42]; when the vaccines arrive, how will they be distributed? How many are there, and where are those people who need it most? Those are some examples where up-to-date socio-economic and demographic data are key (e.g., [85]). However, the last census in Brazil was in 2010, showing the urgency of an EO-based data ecosystem.



**Figure 11.** Using OSM pathways, DTM and social-demographic data from Census 2010, it was possible to estimate that more than 24,000 people, depending on public health services, live more than a 15 min walk away of a Basic Health Care Unity and more than 35,000 are more than 15 min (within 16,000 are more than 30 min) walk distance to a 24 h Health Care Unity (adapted from [86]).

#### 4.2. EO Data for Local and National SDG 11 Monitoring

The first indicator of the “urban” SDG 11 acknowledges the importance of ensuring “access for all to adequate, safe and affordable housing and basic services and upgrade slums” (SDG 11 Target). However, data to support this target is, at the local level, mostly absent [87]. Commonly, SDG 11 statistics are reported as country-level estimates without the inclusion of local spatial data. Local key stakeholders (e.g., city authorities and communities) do not have easy access to ARD or summative data on their communities [11]. There

are many relevant data available to capture the geography of deprivation and would allow local experts to set thresholds in data models (e.g., such as the proposed data ecosystem by IDEAMAPS). For example, such data could support defining and prioritising slums for upgrading. However, the presently available data on most cities are patchy, inconsistent and simply not accessible for non-EO or geo-spatial experts. A new initiative, where the authors of this paper are actively part of, is the development of an SDG 11 toolkit [88] that guides local users in accessing EO-based data relevant to local and national SDG 11 reporting. An important element of this toolkit is the offering of capacity building in using such data. The inclusion of communities is also very important to account for the data ethics in terms of empowering access to data and derived information. Furthermore, it is essential to provide easy access for local communities to data about their location and to support their interaction with data systems, e.g., to decide which data should be included and which not. As such, it is also important to cover not only information on what is missing in communities but also about their assets.

#### *4.3. Data Dissemination, Validation and Accounting for Uncertainties*

Data dissemination needs to respond to several key questions. Such key questions relate to (i) how have data been produced? (clear documentation); (ii) how have data been validated? and (iii) what are the underlying uncertainties in data? As such, it is essential to quantify uncertainties in the produced maps [70]. To be of true value and gain trust, any map or (socio-economic or demographic) variable prediction should be associated with an uncertainty level or confidence interval. Uncertainties should be reported as an overall summary for the entire dataset and spatial information associated with a specific geographic location (e.g., a grid cell in the case of the gridded system). Thus, the three elements of uncertainties need to be reported along with data (i) uncertainties in data (e.g., input data used as input for training the model) usually called aleatoric uncertainty; (ii) uncertainties in the prediction of the model, usually called epistemic uncertainty and (iii) spatial uncertainty (e.g., related to the general geography of an area). As EO experts, we need to understand when aiming to support evidence-based policymaking and to promote data-driven policy- and decision-making that the quality of data products needs to be carefully assessed as well as communicated in a way that non-EO experts can understand.

In addition to uncertainty quantification, there is a growing interest in making machine learning and deep learning models more interpretable and understandable, aiming at neural networks that provide understandable justifications for their output, leading to insights about the inner workings [89,90]. A clear response to the black box problem of AI algorithms and the reduction in the technical barrier. For example, in several AI-based outputs, systematic biases can be observed [91] but are difficult to explain, which endangers the easy interactions with users and trust in data.

## **5. Conclusions**

Socio-economic transformations are rapid in cities with large demographic dynamics. Available census-based socio-economic and demographic data are rapidly outdated in such cities. This leaves stakeholders, ranging from community groups to NSO, with insufficient bases for informed decision-making. Furthermore, the digital divide makes deprived urban areas less covered in most datasets used for decision-making processes. High uncertainties about data on deprived urban areas exist; such uncertainties are reflected in population data, local socio-economic data as well as data that flow into SDG 11 statistics. AI-based methods combined with EO data have great potential to fill such data gaps. However, low cost and no-cost EO data are important to allow the scalability of methods. To increase the reach and standardization of these techniques and data sources flexible, multi-scale and gridded mapping systems are required. Such systems allow combining relevant dimensions of socio-economic conditions with demographic estimates in the form of data cubes. For the provision of locally relevant and acceptable data ecosystems, the protection of privacy needs to be well addressed. While the potential of EO (data and methods) to

supplement official statistics has been demonstrated, solutions are required in the areas of communication and engagement; they are key to promoting collaboration and thus progress towards sustainability.

**Author Contributions:** Conceptualisation, M.K., P.M.G., O.J.J.C.; methodology, M.K., P.M.G., O.J.J.C., E.V.G.; D.R.T., S.V., Á.A., I.O., M.N., P.L.B., C.P., J.L.O.Q.; formal analysis, M.K., P.M.G., O.J.J.C., E.V.G.; D.R.T., S.V., Á.A., P.L.B., C.P., J.L.O.Q.; writing—original draft preparation, M.K., P.M.G., O.J.J.C.; and writing—review and editing, M.K., P.M.G., O.J.J.C., E.V.G.; D.R.T., S.V., Á.A., I.O., M.N., P.L.B., C.P., J.L.O.Q.. All authors have read and agreed to the published version of the manuscript.

**Funding:** The research pertaining to these results received financial aid from the Belgian Federal Science Policy according to the agreement of subsidy no. (SR/11/380) (SLUMAP: <http://slumap.ulb.be/> accessed on 10 June 2021), from NWO grant number VI. Veni. 194.025 and from the GCRF Digital Innovation for Development in Africa panel (EPSRC Reference: EP/T029900/1).

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A. The Global COVID-19 Survey of National Statistical Offices

Official survey operations have impacted 40% of all countries that responded. However, results show that statistical agencies are adapting their survey operations by using alternative data collection methods: 58% of countries are relying on phone surveys; 53%, on administrative data; and 34% on web surveys. In most cases, all three alternative sources are used across income levels. In general, administrative data production is less affected. Here, 56% of offices reported moderate or severe impediments (mostly high-income countries).

**Table A1.** Results of COVID-19 survey (Source: Monitoring the state of statistical operations under the COVID -19 Pandemic. The World Bank and United Nations Statistical Division. Available at: <https://covid-19-response.unstatshub.org/statistical-programmes/covid19-nso-survey/> (accessed on 20 August 2021)).

Type of Census Planned for 2020	Number of Countries That Were Planning One	Saw an Impact on Preparatory Activities (Percentage of Those Who Answered)	Had to Postpone Field Work to Later in 2020 or to 2021 or Beyond (Percentage of Those Who Answered)
Population and Housing Census	61	58%	53%
Agricultural Census	44	50%	55%
Business Census	26	57%	64%

In general, two-thirds of NSOs are impacted to produce essential statistics and to meet international reporting requirements. However, low-income countries are more impacted. All low-income countries responding to the survey confirmed that COVID-19 affected international reporting requirements. While 48% of high-income countries reported that COVID-19 did not impact international reporting.

Furthermore, it is not only the production of statistical information that is being currently challenged: geo-spatial information is also heavily impacted by the pandemic. The National Institute of Statistics and Geography (INEGI) of Mexico did a survey on the actions implemented by both NSOs and national geographic institutes (NGIs) due to COVID-19; some regional or global organisms relevant to these activities were also included in the surveys.

Regarding the NGIs, 24 countries answered the survey (Argentina, Bolivia, Brazil, Canada, Chile, Colombia, Costa Rica, Ecuador, El Salvador, Spain, Finland, France, Guatemala, Honduras, Indonesia, Italy, Norway, Panama, Paraguay, Peru, Dominican Republic's Military Cartographic Institute, United Kingdom, Uruguay and Venezuela)

along with the Pan American Institute for Geography and History (PAIGH). Other 9 did not publish the information (Germany, Cuba, Denmark, Japan, Nicaragua, the Dominican Republic's National Geographic Institute José Joaquín Hungary Morell, Russia, Sweden and USA). From the 25 institutions that answered, 12 (Bolivia, Colombia, El Salvador, Finland, Guatemala, Honduras, Panama, Paraguay, Dominican Republic's Military Cartographic Institute, United Kingdom, Uruguay and Venezuela) mention having taken measures of confinement and adopting a home office modality; 11 (Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Spain, Finland, France, Italy and United Kingdom) mention having changed the face-to-face modality to digital (through servers and internet sites) in customer service, consultation and purchase of products and services and procedures; 5 (Argentina, Colombia, Ecuador, Italy and Peru) mention having suspended events, training courses and cultural activities (planetary museums); and 3 (Ecuador, Peru and PAIGH) mentioned having suspended meetings and work trips. None of the informants reported information on conducting surveys or fieldwork.

Regarding the statistical institutions, 32 NSOs answered the survey (Argentina, Germany, Australia, Austria, Belgium, Brazil, Canada, Korea, Colombia, Costa Rica, Denmark, Ecuador, Spain, 2 from the USA, Estonia, France, Finland, India, Ireland, Italy, Japan, Luxembourg, Norway Netherlands, Paraguay, Peru, Poland, Dominican Republic, Sweden, Switzerland and Uruguay), along with UNECE and UNSD. Of the 34 responding institutions, 24 reported adopting remote work modality (total or partial); 8 reported having suspended fieldwork; 18 reported having suspended some of their surveys (those carried out face to face); likewise, from these 18 that reported suspension of surveys, 6 (Argentina, Australia, Korea, Denmark, Ireland and Poland) reported using administrative records to cover for the required information. On the other hand, as an alternative to face-to-face surveys, 18 entities reported having switched to the telephone or online survey modality; and 10 reported having digital services for data capture and service provision. Furthermore, Ireland and Sweden mentioned plans to modify or include some questions in their current surveys to measure the changes or effects of the pandemic on the labour force.

## References

1. UN-Habitat. Cities Alliance. In *Analytical Perspective of Pro-Poor Slum Upgrading Frameworks*; UN-HABITAT: Nairobi, Kenya, 2006.
2. UN-DESA. *The Sustainable Development Goals Report 2018*; United Nations: New York, NY, USA, 2018.
3. Aguilar, R.; Kuffer, M. Cloud Computation Using High-Resolution Images for Improving the SDG Indicator on Open Spaces. *Remote Sens.* **2020**, *12*, 1144. [CrossRef]
4. Kohli, D.; Sliuzas, R.V.; Kerle, N.; Stein, A. An ontology of slums for image-based classification. *Comput. Environ. Urban Syst.* **2012**, *36*, 154–163. [CrossRef]
5. UN-Habitat. Metadata Indicator 11.1.1. 2018. Available online: [https://unhabitat.org/sites/default/files/2020/06/metadata\\_on\\_sdg\\_indicator\\_11.1.1.pdf](https://unhabitat.org/sites/default/files/2020/06/metadata_on_sdg_indicator_11.1.1.pdf) (accessed on 20 August 2021).
6. Openshaw, S. *The Modifiable Areal Unit Problem*; Geobooks: Norwich, UK, 1984.
7. Thomson, D.R.; Linard, C.; Vanhuyse, S.; Steele, J.E.; Shimoni, M.; Siri, J.; Caiaffa, W.T.; Rosenberg, M.; Wolff, E.; Grippa, T.; et al. Extending Data for Urban Health Decision-Making: A Menu of New and Potential Neighborhood-Level Health Determinants Datasets in LMICs. *J. Urban Health* **2019**, *96*, 514–536. [CrossRef]
8. Abascal, Á.; Rothwell, N.; Shonowo, A.; Thomson, D.R.; Elias, P.; Else, H.; Yeboah, G.; Kuffer, M. "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* **2021**, 2021020242. [CrossRef]
9. United Nations. The Global COVID-19 Survey of National Statistical Offices. Available online: <https://unstats.un.org/unsd/covid19-response/covid19-nso-survey-report.pdf> (accessed on 18 November 2020).
10. Castelán, C.R.; Weber, I.; Jacques, D.; Monroe, T. Making a better poverty map. In *World Bank Blogs*; World Bank: Washington, DC, USA, 2019.
11. Thomson, D.R.; Kuffer, M.; Boo, G.; Hati, B.; Grippa, T.; Else, H.; Linard, C.; Mahabir, R.; Kyobutungi, C.; Maviti, J.; et al. Need for an Integrated Deprived Area "Slum" Mapping System (IDEAMAPS) in Low- and Middle-Income Countries (LMICs). *Soc. Sci.* **2020**, *9*, 80. [CrossRef]
12. Thomson, D.R.; Gaughan, A.E.; Stevens, F.R.; Yetman, G.; Elias, P.; Chen, R. Evaluating the Accuracy of Gridded Population Estimates in Slums: A Case Study in Nigeria and Kenya. *Urban Sci.* **2021**, *5*, 48. [CrossRef]
13. Lilford, R.; Kyobutungi, C.; Ndugwa, R.; Sartori, J.; Watson, S.I.; Sliuzas, R.; Kuffer, M.; Hofer, T.; Porto de Albuquerque, J.; Ezech, A. Because space matters: Conceptual framework to help distinguish slum from non-slum urban areas. *BMJ Glob. Health* **2019**, *4*, e001267. [CrossRef]

14. Carr-Hill, R. Missing millions and measuring development progress. *World Dev.* **2013**, *46*, 30–44. [CrossRef]
15. Kuffer, M.; Wang, J.; Nagenborg, M.; Pfeffer, K.; Kohli, D.; Sliuzas, R.; Persello, C. The Scope of Earth-Observation to Improve the Consistency of the SDG Slum Indicator. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 428. [CrossRef]
16. Wardrop, N.A.; Jochem, W.C.; Bird, T.J.; Chamberlain, H.R.; Clarke, D.; Kerr, D.; Bengtsson, L.; Juran, S.; Seaman, V.; Tatem, A.J. Spatially disaggregated population estimates in the absence of national population and housing census data. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 3529–3537. [CrossRef]
17. Prakash, M.; Ramage, S.; Kavvada, A.; Goodman, S. Open Earth Observations for Sustainable Urban Development. *Remote Sens.* **2020**, *12*, 1646. [CrossRef]
18. Kuffer, M.; Thomson, D.R.; Boo, G.; Mahabir, R.; Grippa, T.; Vanhuyse, S.; Engstrom, R.; Ndugwa, R.; Makau, J.; Darin, E.; et al. The Role of Earth Observation in an Integrated Deprived Area Mapping “System” for Low-to-Middle Income Countries. *Remote Sens.* **2020**, *12*, 982. [CrossRef]
19. Morrison, J. An Introduction to Satellite Imagery and Machine Learning—Azavea Blog. 2019. Available online: <https://www.azavea.com/blog/2019/11/05/an-introduction-to-satellite-imagery-and-machine-learning/> (accessed on 27 February 2021).
20. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [CrossRef]
21. Bergado, J.R.; Persello, C.; Stein, A. Recurrent Multiresolution Convolutional Networks for VHR Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 6361–6374. [CrossRef]
22. Zhang, Z.; Wang, H.; Xu, F.; Jin, Y. Complex-Valued Convolutional Neural Network and Its Application in Polarimetric SAR Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 7177–7188. [CrossRef]
23. Mullissa, A.G.; Persello, C.; Stein, A. PolSARNet: A Deep Fully Convolutional Network for Polarimetric SAR Image Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 5300–5309. [CrossRef]
24. Verma, D.; Jana, A.; Ramamritham, K. Transfer learning approach to map urban slums using high and medium resolution satellite imagery. *Habitat Int.* **2019**, *88*, 101981. [CrossRef]
25. Ajami, A.; Kuffer, M.; Persello, C.; Pfeffer, K. Identifying a Slums’ Degree of Deprivation from VHR Images Using Convolutional Neural Networks. *Remote Sens.* **2019**, *11*, 1282. [CrossRef]
26. Roy, D.; Bernal, D.; Lees, M. An exploratory factor analysis model for slum severity index in Mexico City. *Urban Stud.* **2020**, *57*, 789–805. [CrossRef]
27. Kuffer, M.; Pfeffer, K.; Sliuzas, R.; Baud, I. Extraction of slum areas from VHR imagery using GLCM variance. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 1830–1840. [CrossRef]
28. Duque, J.C.; Patino, J.E.; Betancourt, A. Exploring the Potential of Machine Learning for Automatic Slum Identification from VHR Imagery. *Remote Sens.* **2017**, *9*, 895. [CrossRef]
29. Wurm, M.; Stark, T.; Zhu, X.X.; Weigand, M.; Taubenböck, H. Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks. *ISPRS J. Photogramm. Remote Sens.* **2019**, *150*, 59–69. [CrossRef]
30. Liu, R.; Kuffer, M.; Persello, C. The Temporal Dynamics of Slums Employing a CNN-Based Change Detection Approach. *Remote Sens.* **2019**, *11*, 2844. [CrossRef]
31. Taylor, L. Safety in numbers? Group privacy and big data analytics in the developing world. In *Group Privacy: New Challenges of Data Technologies*; Taylor, L., Floridi, L., van der Sloot, B., Eds.; Springer: Dordrecht, The Netherlands, 2017; pp. 13–36.
32. Arora, P. General Data Protection Regulation—A Global Standard? Privacy Futures, Digital Activism, and Surveillance Cultures in the Global South. *Surveill. Soc.* **2019**, *17*, 717–725. [CrossRef]
33. Beukes, A. *Making the Invisible Visible: Generating Data on ‘Slums’ at Local, City and Global Scales*; International Institute for Environment and Development: London, UK, 2015.
34. Gevaert, C.M.; Sliuzas, R.; Persello, C.; Vosselman, G. Evaluating the Societal Impact of Using Drones to Support Urban Upgrading Projects. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 91. [CrossRef]
35. Mahabir, R.; Croitoru, A.; Crooks, A.; Agouris, P.; Stefanidis, A. A critical review of high and very high-resolution remote sensing approaches for detecting and mapping slums: Trends, challenges and emerging opportunities. *Urban Sci.* **2018**, *2*, 8. [CrossRef]
36. Mahabir, R.; Agouris, P.; Stefanidis, A.; Croitoru, A.; Crooks, A.T. Detecting and mapping slums using open data: A case study in Kenya. *Int. J. Digit. Earth* **2018**, 1–25. [CrossRef]
37. Mahabir, R.; Crooks, A.; Croitoru, A.; Agouris, P. The study of slums as social and physical constructs: Challenges and emerging research opportunities. *Reg. Stud. Reg. Sci.* **2016**, *3*, 399–419. [CrossRef]
38. Rangelova, E.; Weel, B.; Roy, D.; Kuffer, M.; Pfeffer, K.; Lees, M. Image based classification of slums, built-up and non-built-up areas in Kalyan and Bangalore, India. *Eur. J. Remote Sens.* **2018**, *52*, 40–61. [CrossRef]
39. SDI. *Strategic Plan 2018–2022*; SDI: Cape Town, South Africa, 2018.
40. SDI. *Know Your City: Slum Dwellers Count*; SDI: Cape Town, South Africa, 2017.
41. Leonita, G.; Kuffer, M.; Sliuzas, R.; Persello, C. Machine Learning-Based Slum Mapping in Support of Slum Upgrading Programs: The Case of Bandung City, Indonesia. *Remote Sens.* **2018**, *10*, 1522. [CrossRef]
42. Brito, P.L.; Kuffer, M.; Koeva, M.; Pedrassoli, J.C.; Wang, J.; Costa, F.; Freitas, A.D.d. The Spatial Dimension of COVID-19: The Potential of Earth Observation Data in Support of Slum Communities with Evidence from Brazil. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 557. [CrossRef]

43. European Forum for Geography and Statistics (EFGS). New Dataset on Statistical Grids. Available online: <https://www.efgs.info/2020/02/27/new-dataset-on-statistical-grids/> (accessed on 25 April 2021).
44. Ansari, R.A.; Buddhiraju, K.M. Textural segmentation of remotely sensed images using multiresolution analysis for slum area identification. *Eur. J. Remote Sens.* **2019**, *52*, 74–88. [[CrossRef](#)]
45. Gadiraju, K.K.; Vatsavai, R.R.; Kaza, N.; Wibbels, E.; Krishna, A. Machine Learning Approaches for Slum Detection Using Very High Resolution Satellite Images. In Proceedings of the 2018 IEEE International Conference on Data Mining Workshops (ICDMW), Singapore, 17–20 November 2018; pp. 1397–1404.
46. Prabhu, R.; Alagu Raja, R.A. Urban Slum Detection Approaches from High-Resolution Satellite Data Using Statistical and Spectral Based Approaches. *J. Ind. Soc. Remote Sens.* **2018**, *46*, 2033–2044. [[CrossRef](#)]
47. Schmitt, A.; Sieg, T.; Wurm, M.; Taubenböck, H. Investigation on the separability of slums by multi-aspect TerraSAR-X dual-co-polarized high resolution spotlight images based on the multi-scale evaluation of local distributions. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *64*, 181–198. [[CrossRef](#)]
48. Kuffer, M.; Pfeffer, K.; Sliuzas, R.; Baud, I.; van Maarseveen, M. Capturing the Diversity of Deprived Areas with Image-Based Features: The Case of Mumbai. *Remote Sens.* **2017**, *9*, 384. [[CrossRef](#)]
49. Gevaert, C.M.; Persello, C.; Elberink, S.O.; Vosselman, G.; Sliuzas, R. Context-Based Filtering of Noisy Labels for Automatic Basemap Updating from UAV Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 2731–2741. [[CrossRef](#)]
50. Li, Y.; Huang, X.; Liu, H. Unsupervised Deep Feature Learning for Urban Village Detection from High-Resolution Remote Sensing Images. *Photogramm. Eng. Remote Sens.* **2017**, *83*, 567–579. [[CrossRef](#)]
51. Shabat, A.; Tapamo, J.-R. A comparative study of the use of local directional pattern for texture-based informal settlement classification. *J. Appl. Res. Technol.* **2019**, *15*, 250–258. [[CrossRef](#)]
52. Gevaert, C.M.; Persello, C.; Sliuzas, R.; Vosselman, G. Informal settlement classification using point-cloud and image-based features from UAV data. *ISPRS J. Photogramm. Remote Sens.* **2017**, *125*, 225–236. [[CrossRef](#)]
53. Mboga, N.O.; Persello, C.; Bergado, J.; Stein, A. Detection of informal settlements from VHR images using Convolutional Neural Networks. *Remote Sens.* **2017**, *9*, 1106. [[CrossRef](#)]
54. Badmos, O.S.; Rienow, A.; Callo-Concha, D.; Greve, K.; Jürgens, C. Urban development in West Africa—Monitoring and intensity analysis of slum growth in Lagos: Linking pattern and process. *Remote Sens.* **2018**, *10*, 1044. [[CrossRef](#)]
55. Wang, J.; Kuffer, M.; Roy, D.; Pfeffer, K. Deprivation pockets through the lens of convolutional neural networks. *Remote Sens. Environ.* **2019**, *234*, 111448. [[CrossRef](#)]
56. Owusu, M.; Kuffer, M.; Belgiu, M.; Grippa, T.; Lennert, M.; Georganos, S.; Vanhuysse, S. Towards user-driven earth observation-based slum mapping. *Comput. Environ. Urban Syst.* **2021**, *89*, 101681. [[CrossRef](#)]
57. Vanhuysse, S.; Georganos, S.; Kuffer, M.; Grippa, T.; Lennert, M.; Wolff, E. Gridded urban deprivation probability from open optical imagery and dual-pol sar data. In Proceedings of the IEEE IGARSS 2021, Brussels, Belgium, 11–16 July 2021.
58. Persello, C.; Kuffer, M. Towards uncovering socio-economic inequalities using VHR satellite images and deep learning. In Proceedings of the IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020; pp. 3747–3750.
59. Williams, T.K.A.; Wei, T.; Zhu, X. Mapping Urban Slum Settlements Using Very High-Resolution Imagery and Land Boundary Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 166–177. [[CrossRef](#)]
60. Prabhu, R.; Parvathavarthini, B.; Alaguraja, A. Integration of deep convolutional neural networks and mathematical morphology-based postclassification framework for urban slum mapping. *J. Appl. Remote Sens.* **2021**, *15*, 014515. [[CrossRef](#)]
61. Upadhyay, V.; Saini, O.; Pandey, K.; Bhardwaj, A. Identification of slum settlements using logistic regression. In Proceedings of the ACRS 2020—41st Asian Conference on Remote Sensing, Deqing, China, 9–11 November 2020.
62. Debray, H.; Kuffer, M.; Persello, C.; Klafus, C.; Pfeffer, K. Detection of Informal Graveyards in Lima using Fully Convolutional Network with VHR Images. In Proceedings of the 2019 Joint Urban Remote Sensing Event (JURSE), Vannes, France, 22–24 May 2019; pp. 1–4.
63. Wang, J.; Kuffer, M.; Pfeffer, K. The role of spatial heterogeneity in detecting urban slums. *Comput. Environ. Urban Syst.* **2019**, *73*, 95–107. [[CrossRef](#)]
64. Engstrom, R.; Harrison, R.; Mann, M.; Fletcher, A. Evaluating the Relationship between Contextual Features Derived from Very High Spatial Resolution Imagery and Urban Attributes: A Case Study in Sri Lanka. In Proceedings of the 2019 Joint Urban Remote Sensing Event (JURSE), Vannes, France, 22–24 May 2019; pp. 1–4.
65. Engstrom, R.; Pavelesku, D.; Tanaka, T.; Wambile, A. Mapping Poverty and Slums Using Multiple Methodologies in Accra, Ghana. In Proceedings of the 2019 Joint Urban Remote Sensing Event (JURSE), Vannes, France, 22–24 May 2019; pp. 1–4.
66. Badmos, O.S.; Rienow, A.; Callo-Concha, D.; Greve, K.; Jürgens, C. Simulating slum growth in Lagos: An integration of rule based and empirical based model. *Comput. Environ. Urban Syst.* **2019**, *77*, 101369. [[CrossRef](#)]
67. Persello, C.; Stein, A. Deep Fully Convolutional Networks for the Detection of Informal Settlements in VHR Images. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 2325–2329. [[CrossRef](#)]
68. Wurm, M.; Taubenböck, H. Detecting social groups from space—Assessment of remote sensing-based mapped morphological slums using income data. *Remote Sens. Lett.* **2018**, *9*, 41–50. [[CrossRef](#)]

69. Ella, L.P.A.; Van Den Bergh, F.; Van Wyk, B.J.; Van Wyk, M.A. A comparison of texture feature algorithms for urban settlement classification. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Boston, MA, USA, 26 September–2 October 2008; Volume 3, pp. III1308–III1311.
70. Pratomo, J.; Kuffer, M.; Martinez, J.; Kohli, D. Coupling uncertainties with accuracy assessment in object-based slum detections, case study: Jakarta, Indonesia. *Remote Sens.* **2017**, *9*, 1164. [CrossRef]
71. Moser, C.O.N. The asset vulnerability framework: Reassessing urban poverty reduction strategies. *World Dev.* **1998**, *26*, 1–19. [CrossRef]
72. Müller, I.; Taubenböck, H.; Kuffer, M.; Wurm, M. Misperceptions of Predominant Slum Locations? Spatial Analysis of Slum Locations in Terms of Topography Based on Earth Observation Data. *Remote Sens.* **2020**, *12*, 2474. [CrossRef]
73. Zhu, Z.; Zhou, Y.; Seto, K.C.; Stokes, E.C.; Deng, C.; Pickett, S.T.A.; Taubenböck, H. Understanding an urbanizing planet: Strategic directions for remote sensing. *Remote Sens. Environ.* **2019**, *228*, 164–182. [CrossRef]
74. Ghebreyesus, T.A. WHO Director-General’s Opening Remarks at the Media Briefing on COVID-19-11 March 2020. Available online: <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19--11-march-2020> (accessed on 10 September 2020).
75. van Deursen, J.A.M. Digital Inequality During a Pandemic: Quantitative Study of Differences in COVID-19-Related Internet Uses and Outcomes Among the General Population. *J. Med. Internet Res.* **2020**, *22*, e20073. [CrossRef]
76. Shadmi, E.; Chen, Y.; Dourado, I.; Faran-Perach, I.; Furler, J.; Hangoma, P.; Hanvoravongchai, P.; Obando, C.; Petrosyan, V.; Rao, K.D.; et al. Health equity and COVID-19: Global perspectives. *Int. J. Equity Health* **2020**, *19*, 1–16. [CrossRef] [PubMed]
77. Holden, M. COVID-19 death rate in deprived areas in England double that of better off places: ONS. *Reuters*, 2020. Available online: <https://www.reuters.com/article/us-health-coronavirus-britain-deprived-idUSKBN22D51O> (accessed on 20 August 2020).
78. Iacobucci, G. Covid-19: Deprived areas have the highest death rates in England and Wales. *BMJ* **2020**, *369*, m1810. [CrossRef] [PubMed]
79. Secretaria Municipal de Saúde/Prefeitura de Salvador. TABNET. Available online: <http://www.tabnet.saude.salvador.ba.gov.br> (accessed on 30 September 2020).
80. Centers for Disease Control and Prevention. Social Distancing. Keep a Safe Distance to Slow the Spread. Available online: <https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/social-distancing.html> (accessed on 30 September 2020).
81. World Health Organization Country Office for Thailand. The 6 Steps. Available online: [https://www.who.int/docs/default-source/searo/thailand/who-tha-six-steps.pdf?sfvrsn=b81cac2b\\_0](https://www.who.int/docs/default-source/searo/thailand/who-tha-six-steps.pdf?sfvrsn=b81cac2b_0) (accessed on 30 September 2020).
82. Barber, S.; Diez Roux, A.V.; Cardoso, L.; Santos, S.; Toste, V.; James, S.; Barreto, S.; Schmidt, M.; Giatti, L.; Chor, D. At the intersection of place, race, and health in Brazil: Residential segregation and cardio-metabolic risk factors in the Brazilian Longitudinal Study of Adult Health (ELSA-Brasil). *Soc. Sci. Med.* **2018**, *199*, 67–76. [CrossRef] [PubMed]
83. Unger, A.; Riley, L.W. Slum Health: From Understanding to Action. *PLoS Med.* **2007**, *4*, e295. [CrossRef]
84. Hagan, J.E.; Moraga, P.; Costa, F.; Capian, N.; Ribeiro, G.S.; Wunder, E.A., Jr.; Felzemburgh, R.D.M.; Reis, R.B.; Nery, N.; Santana, F.S.; et al. Spatiotemporal Determinants of Urban Leptospirosis Transmission: Four-Year Prospective Cohort Study of Slum Residents in Brazil. *PLoS Negl. Trop. Dis.* **2016**, *10*, e0004275. [CrossRef]
85. Johns Hopkins Resource Centre. Coronavirus Resource Center. Available online: <https://coronavirus.jhu.edu/map.html> (accessed on 30 September 2020).
86. Brito, P.L.; Viana, M.S.; Delgado, J.P.M.; Brandão, A.C.; Pedrassoli, J.C.; Pedreira Júnior, J.U.; Souza, F.A. *Nota Técnica 04—Alertas e Propostas de Ações para Península de Itapagipe Baseadas em Análises Geoespaciais de Suporte ao Combate à COVID-19*; GeoCombate: Salvador, Brazil, 2020.
87. Kuffer, M.; Pfeffer, K.; Sliuzas, R.; Taubenböck, H.; Baud, I.; Maarseveen, M.v. Capturing the Urban Divide in Nighttime Light Images From the International Space Station. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 2578–2586. [CrossRef]
88. UN-Habitat. The Earth Observations Toolkit for Sustainable Cities and Human Settlements. Available online: <https://eo-toolkit-geo-un-habitat.opendata.arcgis.com> (accessed on 15 March 2021).
89. Samek, W.; Montavon, G.; Lapuschkin, S.; Anders, C.J.; Müller, K.-R. Explaining Deep Neural Networks and Beyond: A Review of Methods and Applications. *Proc. IEEE* **2021**, *109*, 247–278. [CrossRef]
90. Campos-Taberner, M.; García-Haro, F.J.; Martínez, B.; Izquierdo-Verdiguier, E.; Atzberger, C.; Camps-Valls, G.; Gilabert, M.A. Understanding deep learning in land use classification based on Sentinel-2 time series. *Sci. Rep.* **2020**, *10*, 17188. [CrossRef]
91. Ntoutsis, E.; Fafalios, P.; Gadiraju, U.; Iosifidis, V.; Nejdli, W.; Vidal, M.-E.; Ruggieri, S.; Turini, F.; Papadopoulos, S.; Krasanakis, E.; et al. Bias in data-driven artificial intelligence systems—An introductory survey. *WIREs Data Min. Knowl. Discov.* **2020**, *10*, e1356. [CrossRef]