PEAK Urban

The need for real time and granular data to study the urban economy

Author: Shriya Anand, Viola Lewis, Herry Gulabani

Indian Institute for Human Settlements

The Government of India is pursuing a growth agenda that is driven by cities. However, to facilitate a post-COVID economic recovery that also prioritises job creation, the central and state governments need frequent and granular data to help guide policy and investment decisions. This brief demonstrates the existing gaps in the urban economic data landscape and suggests datasets (i.e. the Economic Census) that offer some potential.

Key insights:

- The Government of India is pursuing an economic growth agenda that is driven by cities, signalling a shift from its earlier policy priorities that focused on rural areas.
- While India has historically built a robust statistical system consisting of regular Censuses as well as sample surveys, there have been disruptions to this over the past decade (Bhattacharya, 2022).
- Further, there are gaps in the ways by which data is released, the frequency of data collection, and the level of spatial granularity available to policymakers, planners, and researchers.
- Issues in the quality of datasets hinder the ability of central, state and city governments to more effectively make economic and urban plans, as well as prioritise areas for infrastructure investments.
- Datasets, such as the Economic Census, offer some potential in this direction; however, there are still gaps that would need to be addressed to utilise this.

What hangs in the balance?

There is an emerging policy consensus around promoting cities as engines of economic growth (XV Finance Commission, 2020, pg. 169). However, India had been facing a slowdown of employment growth even before the pandemic, and job creation is not keeping pace with the increase in the labour force (Kanan and Raveendran, 2019). This employment crisis has been further intensified due to the pandemic (Azim Premji University, 2021). Therefore, we argue that it is important to focus on a post-COVID growth agenda that also prioritises job creation.

In order to do this, planners and policymakers need a data environment that provides information on multiple aspects of urban economies and labour markets. Currently, data on the economies of cities is fragmented across multiple data sources (summarised in Table 1) and has issues in the ways it is released, including interoperability, frequency and spatial granularity (summarised in Box 2). There is a set of sources that capture data on firms and a separate set of surveys on employment. Since the vast majority (70-80%) of the urban workforce in India works in the informal economy and as self-employed workers, there is a need to capture labour market data from employment surveys in addition to data on formal firms. Further, these datasets are often not released along with spatial identifiers, making it difficult to understand either firm geographies or labour geographies of the city, which limits the ability of planners and policymakers to prioritise investments in areas like infrastructure.

In this brief, we summarise the data landscape, including a review of the different sources that can be used to understand the Indian urban economy (Table 1). We also undertake a mapping of the way other countries use data on urban economies for planning (see Box 3). We then present our work on using the Economic Census data for mapping Bangalore's economic and employment geography and conclude with a set of limitations as well as possibilities for moving forward.

Datasets currently used to study urban economies

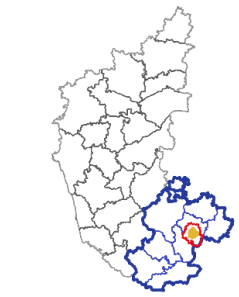
Economic data in India is available in the form of administrative data, censuses, sample surveys and private data sets. Data sources such as the Annual Survey of Industries (ASI), the Ministry of Corporate Affairs (MCA), Centre for Monitoring Indian Economy (CMIE) and the Economic Census (EC) have information on firms, the type of ownership, industry sector, indicators on firm size, among others. The National Sample Survey Organisation (NSSO) carries out household surveys that focus on various aspects of employment, such as the labour force participation rate, unemployment rate, worker population ratio, and the industrial distribution and occupational distribution of workers. Some of these datasets cover both formal and informal jobs while some others focus on informal jobs only. For example, the NSSO Employment Unemployment survey (which enumerates employment from the labour supply, or household, side) and the EC (which enumerates employment from the labour demand, or firm, side) include both formal and informal employment and firms across all jobs sectors, while the ASI covers only registered manufacturing and the NSSO Unorganized Manufacturing survey covers informal manufacturing.

The spatial granularity at which data is available for each of these data sources is varied, too. The EC is released at the highest level of spatial granularity, the Enumeration Block (EB) level (See Box 1). The ASI, MCA and CMIE collect data on firms that includes address information only of firm headquarters and not plant locations, and they release unit-level data on firms that can be mapped to a district, while the NSSO household surveys on employment are representative at the NSS region level. The NSS region consists of several districts, and is too coarse to carry out any meaningful urban-scale analysis of employment. We return to this point in our discussion of the Economic Census as a possible source of information.

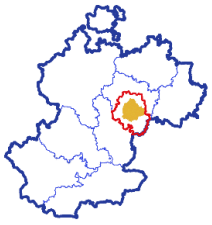
Understanding the different scales of spatial units for analysis

India is divided into states, which are further sub-divided into administrative divisions called districts. Districts consist of sub-districts or taluks, within which there are rural settlements (villages) and urban settlements (towns and cities). According to the 2011 Census, the state of Karnataka has 30 districts including Bangalore Urban district (see Image 1 below). The NSSO's sampling frame uses groups of districts, termed NSS regions, that are grouped together based on various indicators including similar geographical features, rural population densities, etc. (MOSPI, 2001). Bangalore Urban district is part of the "Inland southern region" which consists of 9 districts (see 2 in Image below). Karnataka state has a population of 61 million, of which 23 million people live in the 9 districts forming the NSS region that contains Bangalore (Census, 2011). Bangalore Urban district consists of 16 towns including Bangalore (BBMP, or Bruhat Bengaluru Mahanagara Palike is the municipality governing Bangalore, and its boundary is indicated in yellow in Image 3) and 588 villages (see Image 3). Bangalore Urban district has a population of 9.6 million, of which 8.5 million people live in Bangalore city, i.e. within BBMP limits (Census,2011).

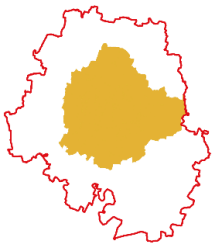
For the purpose of urban analysis or policymaker, the sub-divisions within cities become significant. The Enumeration Block (EB) is the basic unit of measurement within urban areas and is created at the time of population enumeration as part of the decennial Census. An EB consists of around 100-125 households with a population of 650-700 persons (MOSPI, 2013). EBs are grouped together to form a larger administrative unit called the Ward. For the city of Bangalore, a population of 8.5 million in 2011 is distributed across 198 wards (see Image 4 below) and over 16,000 EBs. Wards in Bangalore typically have a population of around 40,000 persons. While the EB is an adequate level at which to analyse datasets such as the Economic Census, EB boundaries are not made available in the public domain, therefore limiting the ability of researchers or policymakers to use this to understand urban economies using this data. We show the EB boundaries for one of the wards (see Image 6 below) released by the Census as a sample.



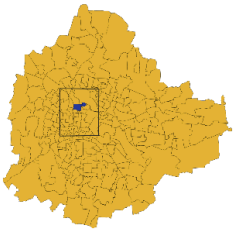
1. Karnataka State with district boundaries



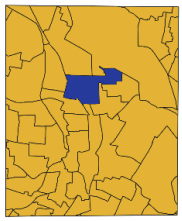
2. NSS Region (9 Districts)



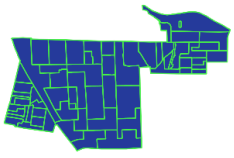
3. Bangalore District and Municipal Boundaries



4. BBMP Ward Boundaries (198 Wards)



5. Sample Representation of Wards



6. EB Boundaries within a ward

Table 1: Datasets currently used to study urban economies
A few key economic data sources, their usefulness and limitations are discussed below.

Dataset	Coverage	Fields depicting business size and sector information	Data Collection Frequency	Smallest Spatial Unit	Ease of working with this data
Annual Survey of industries (ASI)	- Registered manufacturing firms - Units with > 100 employees are enumerated; the rest are covered under a separate sample survey carried out by the ASI	- Output, Value Added - Employee Count - Sector (Rural/Urban) - NIC Code	- Annual	Unit level data is released with only district as the spatial indicator	- Allows tracking of changes within formal manufacturing over time, at national, state and district levels
National Sample Survey Organisation (NSSO) - Unorganised Manufacturing Survey	- Covers unregistered manufacturing - Own-account enterprises, directory manufacturing establishments (DME) employing six or more workers and non-directory manufacturing establishments (NDME) employing less than six workers are covered by this survey	- Information on employment, assets owned, loans, expenses, etc.	- Done at 5-year interval in the past but not since 2006	Sample is representative at the NSSO region level (a set of districts)	- Unit level data on firms is released, with district as the spatial identifier
NSSO - Employment Unemployment Survey	- Employment based on self-reporting from households - Last survey was in 2011-12, and has been replaced by the Periodic Labour Force Survey (PLFS) conducted annually since 2017	- Employment share per sector at region	- Used to be carried out every 5 years, but now is annual	Sample is representative at the NSSO region level (a set of districts)	- Unit level data on households is released - Considered the most authoritative source on employment measurement in India - However, the samples are not thick enough to analyse the labour markets of cities
Economic Census	- Complete enumeration of all organised and 'unorganised' firms excluding crop plantation, public administration, defence and compulsory social security - One anonymised record per enumerated firm	- Number of employees per firm - NIC code as sector	Earlier rounds at 10 year intervals, later rounds are in 6-8 year intervals	Enumeration Block	- Has identifiers like ward and EB code; hence interoperable with Population Census - Employment is found to be undercounted
Population Census	- Complete enumeration of all households	- Total employment by gender aggregated at the ward level - Employment by NIC Code is released at the district level - Employment by occupational code is also captured, and released at the district level	Every 10 years	Employment by sector at the district level	- The Population Census is the most comprehensive coverage of households - It allows industrial and occupational analysis at the city scale; however, this cannot be carried out at a sub-city scale - At the ward level, only aggregate workforce numbers are available, and can be disaggregated by gender - Analysis by sector, or occupational classification, is not possible at the ward level
Ministry of Corporate Affairs (MCA)	- Any company/LLP registered with Registrar of Companies	- Authorised Capital - Paid Up Capital	Monthly	Company name and registered address	- Has company name and registered office address for identifiers - Registered address and location of operation or plant location may be different; therefore, this database cannot be used to understand spatial concentration of employment
Centre for Monitoring Indian Economy (CMIE) ProwessDX	- A firm database that includes all publicly listed companies	- Contains all information on firms that is released in their annual statement, including NIC code, employment, etc.	Quarterly	Company name and registered address	- Paid data - Registered address and location of operation or plant location may be different, therefore this database cannot be used to understand the spatial concentration of employment
Micro Small and Medium Enterprises (MSME) Census	- Enumeration for registered MSMEs and sample survey for the unregistered - MSME 2006-07 is the last enumerated data on MSMEs	- Output - Value Added - Employee Count - NIC code	Not enumerated anymore	Postal Code, Village/Town code	- Postal code of the location of the enterprise and district code are present - A direct comparison between MSME 2000 and MSME 2006 at the city level is a challenge because of the lack of a common spatial identifier between the two datasets

Limitations of these datasets

Lack of accurate spatial locations for formal firms: Even though a considerable amount of data is available for formal firms, accurate spatial locations for these are lacking. Company data released by the Ministry of Corporate Affairs (MCA) and private data sources such as the Centre for Monitoring Indian Economy (CMIE) have only the registered addresses of firms, which may not capture their multiple offices or plants, like in the case of large firms. Location-based policies and investments in infrastructure would need accurate information on where production is taking place, or where office locations might be concentrated—neither of which are accurately captured by these datasets. Better data on firms and firm locations with separate address data for headquarters, offices and plant locations can easily be made available by adding these specific instructions to the way the firms report their locations.

Lack of spatial granularity for employment data: Employment in formal firms is only a small fraction of the urban labour market (less than 20%). Therefore, it is essential to also map informal employment or labour geographies of the city, and employment data is most systematically measured through self-reporting of labour supply, as captured in household surveys such as the NSS and the PLFS. However, NSSO data is useful to understand the distribution of socio-economic variables and the distribution of employment at the national and state level. Its sampling strategy does not allow for granularity and representation beyond the NSS region level. Therefore, urban scale analysis of labour markets is not possible using NSSO data. Similarly, the Population Census enumerated at the Enumeration Block (EB) level, is released to the public only at the district or ward level and this has limited value for decision-making in the urban context.

Alternate data sources and challenges: Because of the difficulties in accessing granular economic data, researchers often use proxies for economic activity such as night-lights, street view imagery, mobile phone data and social media data. These alternate data sources may not be publicly available, may include privacy concerns, or may not be reliable for studying the urban economy. For example, using night-lights data as a proxy for economic activity has its drawbacks. Night-lights data does not distinguish between value added in different sectors; it cannot differentiate between lights emitted from skilled workforces in corporate offices or from unskilled labour in different industries and domestic habitations (Singhal et al., 2020). Further, night-lights data can underrepresent economic activity in regions with high population density in the form of taller buildings (Bhandari and Roychowdhury, 2011).

International comparison cases

New York: New York City's Department of Planning uses tax-lot level data from the census and quarterly employment survey from the federal Bureau of Labour Statistics to produce knowledge products ranging from macro-level economic statistics (like GDP, city-level employment, sector performance and geography of jobs), to micro-level neighbourhood dependency on various sectors and availability of local economic opportunities. These knowledge products then feed into the city's economic plan and policy initiatives, which consist of sector- and neighbourhood-specific initiatives and tie them with land use, housing and transportation planning to help improve job accessibility for its citizens (City of New York, 2020). Further, the availability and frequency of granular data also help the city to build a risk and resilience strategy for local economies. For example, by comparing survey results post-COVID with pre-COVID data, the city identified neighbourhoods most affected and built recovery strategies (City of New York, 2020).

London: London has a city data analytics programme as part of the Greater London Authority, which provides data expertise and technical support for city planning. The team uses micro-level data from the census to classify London neighbourhoods based on employment and output into different economic typologies. Planners use this data to understand the local demography of people living in the city and develop specific targeted strategies. The master plan uses datasets, such as satellite-generated night-time lights data, to study commercial activity and changes in industrial land use. The analysis provides a better understanding of the economic importance of local areas giving insight into industrial demand and the location of parks. 'Real time' data is also used through the analysis of consumer data, crime records and digital data recorded through transport mobilities. By mapping social deprivation and identifying areas of regeneration the city identifies localities for targeted fund programs like the Regeneration Fund, Enterprise Fund and Growing Places Fund, which encourage new growth and job opportunities (Mayor of London, 2021).

Johannesburg: The Economic Development Policy and Planning (EDPP) directorate develops databases and analyses related to the local economy and its performance for the city of Johannesburg. While the Spatial Development Framework 2040 for Johannesburg analyses macro-economic trends in the city, it also addresses the importance of the spatial distribution of the urban economy in the city's growth. The plan identifies areas with a high concentration of economic activity in city cores and spatial disparity of east-west economic growth to design specific spatial interventions that address these economic inequalities (City of Johannesburg, 2016).

São Paulo: The city of São Paulo identifies local areas with high unemployment rates and with a disconnect between residential areas and economic opportunities. Its plan aims to provide urban and fiscal incentives in such areas- by creating SEZs, specialised parks, etc. and integration them with the city's transportation infrastructure- for creating a balanced distribution of employment opportunities in the city (City of São Paulo, 2014). The state of São Paulo also has an investment Promotion Agency (InvestSP) that uses a GIS-based system consisting of demographic and economic data points to help investors identify locations for their investment (ESRI, 2022).

Master planning in the Indian context is restricted to macro-sector level trends and the regulation of land use and buildings and is not sufficiently integrated with economic planning (Idiculla, 2021). If Indian cities need to be the focus of our economic growth, we need more localised planning which can bridge the gap between central economic planning and city-level economic growth patterns. Indian cities have high intra-city diversity in terms of economy, built form and demographics, which demand more localised solutions. In order to enable this, we need better data to represent these diverse neighbourhood characteristics and prepare targeted policies which address local economies. The above examples, from a very different set of contexts, highlight the ways in which India can improve its data environment to support better economic and urban planning for its cities.

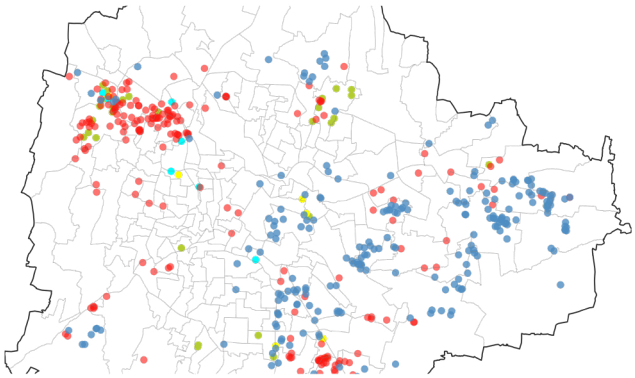
A case for using the Economic Census for mapping Bangalore's economy

Spatial mapping of economic activity for Bangalore, India using the Economic Census (EC)

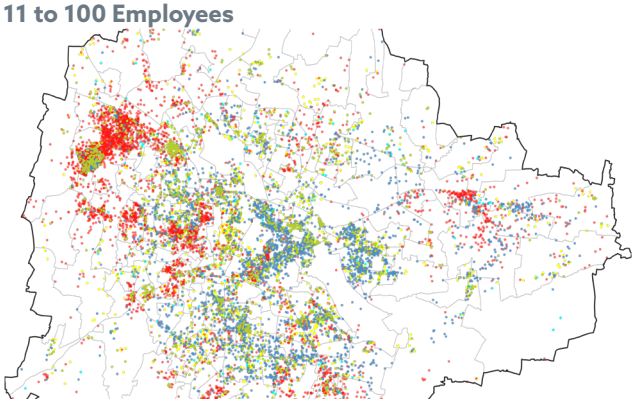
The EC is a complete enumeration of establishments in all sectors excluding crop plantation, public administration, defence and compulsory social security. It is the only source of data on both organized and 'unorganised' or informal sector firms. It uses the EB as the primary geographical unit for data collection and data is also released at the EB level. Using this data along with the EB boundaries, we have created a spatial mapping of economic activity in Bangalore.

Analysis by Size

Very Large (>500 Employees)



11 to 100 Employees



• Manufacturing, • Professional Services, IT, Financial Services, • Food & Accomodation, Admin and other services, • Trade, • Auto trade/repair

Image: Economic activity highlighting size and sector geographically mapped at a fine granularity within Bangalore. Firms are randomly distributed within the enumeration block. The size of the bubble is proportional to the number of employees in the firm.

Key findings from the study:

- According to our analysis at the city level, the EC shows signs of both undercounting and incompleteness when compared with aggregate employment figures from the Population Census. Sources point to undercounting in the EC, especially for very small firms (Asher et al., 2018), but in our analysis of the EC at the city level, we see that very large firms (>100 employees), especially information technology (IT) offices within large IT parks and very large manufacturing enterprises (especially readymade garments) are undercounted. The inability of EC enumerators to access these large gated complexes could be one of the reasons for the missing large firms. Gaps in data for large IT and garments manufacturing were filled in using an alternate source, the Common Business Register, which has a better coverage of large firms compared to the EC.
- A clear distinction between manufacturing in the west and professional services in the east of the city is visible. We also see a pattern of peripheralisation of large firms relative to small and informal firms.

Conclusion and Recommendations

The Economic Census (EC) allows us to map both the formal and informal firms of all sizes and sectors within the city. The quality of this data, however, is problematic, especially since it was delinked from the Population Census in 1998. The previous rounds of EC in 1980 and 1990 were conducted along with the Population Census, with a statutory backing and relatively better-qualified officials. Post-1998, the responsibility of conducting the EC rests with state governments, which leads to concerns such as data reliability and inter-state variation (Raveendran and Unni, 2006). Even in Karnataka, one of the states with relatively better data, substantial effort was required to clean and enhance the data to ensure completeness.

The EC data cannot be mapped for other cities because the government does not release spatial boundaries. To allow urban planners to conduct such spatial analyses across more cities, spatial boundaries at the EB level should be made available.

Beyond mapping firms, there is also a need to improve the spatial resolution of employment data at an intra-city scale, which allows us to also better understand labour geographies in a context of widespread informality.

Data collected by any government entity should be made available in a timely fashion, where needed, with appropriate safeguards, accuracy, high geographic granularity and data interoperability in place (World Bank, 2021). In order to do so, the government can set up a separate statistical regulatory authority that checks for quality, anonymisation and timely releases (Bhattacharya, 2022).

Paper-based government registers should be digitised, and secure connections or APIs should be made available for research agencies to access these.

Once accessed, data needs to be easy to understand with appropriate metadata describing the characteristics of each field and information on how to process them.

References:

Asher S., Murgai R. and Roy, S. (2018). The Geography of Jobs: Explaining Employment Growth in India (2005-2013). World Bank.

Azim Premji University (2021) State of Working India 2021: One year of Covid-19, Centre for Sustainable Employment, Azim Premji University.

https://cse.azimpremjiuniversity.edu.in/wp-content/uploads/2021/08/SWI2021_August_WEB.pdf

Bhattacharya, P. 2022. Make our statistics credible again. Mint.

https://www.livemint.com/opinion/columns/lets-revamp-our-data-ecosystem-a-wish-list-for-2022-11641231517059.html

City of Johannesburg. (2016). Spatial Development Framework 2040.

https://unhabitat.org/sites/default/files/download-manager-files/SDF%20JOHANNESBURG.pdf

City of New York. (2020). Retail activity in NYC: COVID recovery across 24 neighbourhoods.

https://www1.nyc.gov/assets/planning/download/pdf/planning-level/housing-economy/retail-activity-nyc-covid-recovery.pdf

City of New York. (2017). New York Works.

https://newyorkworks.cityofnewyork.us/wp-content/uploads/2020/06/NewYorkWorks-1.pdf

City of Sao Paulo. (2014). Strategic Master Plan.

https://web.archive.org/web/20210115033423/https:/gestaourbana.prefeitura.sp.gov.br/wp-content/uploads/2015/02/Master_plan_english_version.pdf

Economics Statistics Division. (2013). Instruction Manual for Sixth Economic Census. Ministry of Statistics and Programme Implementation.

https://ecostat.telangana.gov.in/PDF/ACT_Manual/Economics_Census_Instruction_Manual.pdf

ESRI. (2022). São Paulo Optimizes Economic Growth with Spatial Intelligence.

https://www.esri.com/about/newsroom/arcnews/sao-paulo-optimizes-economic-growth-with-spatial-intelligence/

Idiculla, M. (2021). Who Plans the Indian city? The anomalies of India's Urban Planning System. The Print.

https://theprint.in/opinion/who-plans-indian-cities-development-authorities-who-still-follow-colonial-masterplans/770913/

Kannan, K.P. and Raveendran, G. (2019). From Jobless to Job-loss growth. Economics and Political Weekly Vol.54, Issue 44.

https://www.epw.in/journal/2019/44/insight/jobless-job-loss-growth.html

Mayor of London. (2021). The London Plan 2021.

https://www.london.gov.uk/sites/default/files/london_plan_2021_presentation_planning_mar2021.pdf

National Sample Survey Organisation. (2003). Concepts and Definitions used in NSS. Ministry of Statistics & Programme Implementation. Government of India.

https://mospi.gov.in/documents/213904/0/concepts_golden.pdf/e98fc072-8660-edd9-f179-ce95674f4ca5?t=1615539414160

Bhandari and Roychowdhury. (2011). Night Lights and Economic Activity in India: A study using DMSP-OLS night time images

https://www.researchgate.net/profile/Laveesh-Bhandari/publication/274133059_Night_Lights_and_Economic_Activity_in_India_A_study_using_DMSP-OLS_night_time_images/links/5b604fa6a6fdccf0b20362c5/Night-Lights-and-Economic-Activity-in-India-A-study-using-DMSP-OLS-night-time-images.pdf

Raveendran, G. and Unni, J. (2006). Are the results of the economic census robust? Economics and Political Weekly, Vol. 41, Issue 33.

https://www.epw.in/journal/2006/33/commentary/are-results-economic-census-robust.html

Singhal et al. (2020). Using night time lights to find regional inequality in India and its relationship with economic development

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0241907#pone.0241907.ref002

World Bank. (2021). Data as a force for public good. World Development Report 2021.

https://wdr2021.worldbank.org/stories/data-as-a-force-for-public-good/

XV Finance Commission. 2020. Finance Commission in COVID times.

https://fincomindia.nic.in/ShowContentOne.aspx?id=9&Section=1

This brief was authored by Shriya Anand, Viola Lewis, and Herry Gulabani. Francisco Obando was the managing editor.

About us

The PEAK Urban programme aims to aid decision-making on urban futures by:

1. Generating new research grounded in the logic of urban complexity;

2. Fostering the next generation of leaders that draw on different perspectives and backgrounds to address the greatest urban challenges of the 21st century;

3. Growing the capacity of cities to understand and plan their own futures.

In PEAK Urban, cities are recognised as complex, evolving systems that are characterised by their propensity for innovation and change. Big data and mathematical models will be combined with insights from the social sciences and humanities to analyse three key arenas of metropolitan intervention: city morphologies (built forms and infrastructures) and resilience; city flux (mobility and dynamics) and technological change; as well as health and wellbeing.

Contact

Shriya Anand
sanand@iihs.ac.in

In PEAK Urban:
peakurban.director@compas.ox.ac.uk

Published by PEAK Urban

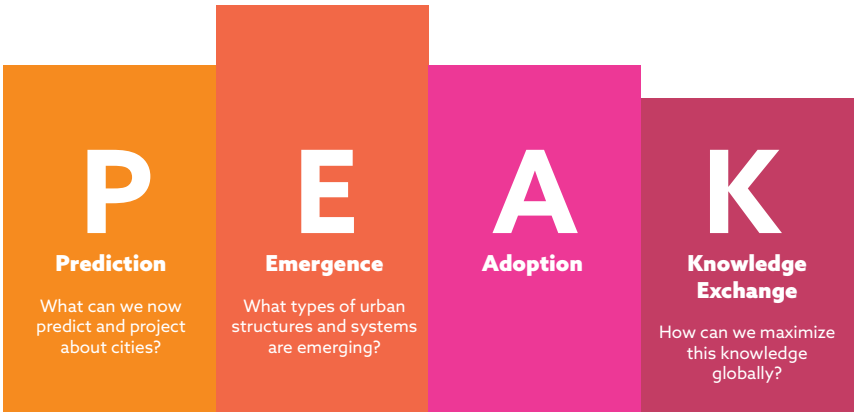
July 2022

PEAK Urban is managed by the Centre on Migration, Policy and Society (COMPAS)

School of Anthropology and Museum Ethnography,
University of Oxford,
8 Banbury Road,
Oxford, OX2 6QS

+44 (0) 1865 274706
@PEAK_Urban
www.peak-urban.org

Our framework



The PEAK Urban programme uses a framework with four inter-related components to guide its work.

First, the sciences of **Prediction** are employed to understand how cities evolve using data from often unconventional sources.

Second, **Emergence** captures the essence of the outcome from the confluence of dynamics, peoples, interests and tools that characterise cities, which lead to change.

Third, **Adoption** signals to the choices made by states, citizens and companies, given the specificities of their places, their resources and the interplay of urban dynamics, resulting in changing local power and influencing dynamics.

Finally, the **Knowledge** component accounts for the way in which knowledge is exchanged or shared and how it shapes the future of the city.

PEAK Urban is funded by UK Research and Innovation as part of the Global Challenges Research Fund.

UK Research and Innovation

GCRF Global Challenges Research Fund

PEAK Urban is a partnership between:

University of Oxford · African Centre for Cities · Peking University 1898 · Universidad EAFIT · iihs Indian Institute for Human Settlements