Precision Mapping of Urban Population Baseline Under the Background of Digital Twin Technology: Innovation Technology Framework and Planning Application Outlook

Wu Jiang, Zhang Yiping, Yuan Ye, Tang Ge, Duan Ruiyan, Xu Shiyan, Kong Ling

Abstract: Digital twin technology is the foundation for intelligent spatial planning. Unlike traditional engineering twins, urban digital twins, due to the dynamic and complex social factors in their objects, have lacked sufficient mapping accuracy regarding population baselines. This has made it difficult for current digital twin technology to provide accurate population information to support urban planning decisions. This paper reviews the development of urban digital twins and their technical challenges, summarizes the current mainstream methods of population mapping and their issues, and, based on the logic of "real space, real population, real demand," proposes an innovative technological framework for the precise mapping of urban population baselines. This framework integrates a wide range of data sources, primarily government data, to build a digital platform for population baseline information that caters to governmental service needs. This technology helps achieve high granularity in urban population analysis, refined community public service provision, and high-quality territorial spatial planning.

Keywords: Urban Digital Twin; Population Baseline; Precision Mapping; Innovation Technology Framework; Planning Application Outlook

China's urbanization has entered a new development stage focused on quality improvement, with new opportunities and challenges for urban development and governance models. One important strategic direction is how to build a perceptible, learnable, well-governed, and adaptive intelligent planning system through digital new spatial ecology. Digital twin technology, as a means to map the real world to the virtual world, is a fundamental tool for achieving smart planning. At the same time, the concept of a "people's city" highlighted in the 20th National Congress of the Communist Party emphasizes that "the people's city is built by the people, and the people's city is for the people." Under the empowerment of digital twin technology, one aspect involves establishing diversified urban digital governance platforms to widely incorporate residents' beneficial thoughts on urban development and renewal, truly realizing "the people's city is built by the people"; another aspect is to precisely map the public service needs of different resident groups through digital mapping, providing an objective basis for spatial public service supply, and truly achieving "the people's city is for the people." However, influenced by traditional engineering twins, the current urban digital twins face the issue of being "people-light," primarily manifesting in insufficient mapping of population information or focusing on mapping only specific types of "populations," leading to significant differences between the "population baseline layer" in the urban digital twin and the real situation. If the urban digital twin cannot accurately map relatively real urban population information, it becomes difficult to identify and evaluate related public service issues accurately during urban planning, construction, and governance, potentially leading to misjudgments about public affairs, causing decision-making errors or failures, and ultimately affecting social equity and residents' well-being. Therefore, accurately mapping the "population baseline layer" in urban digital twins is crucial. Based on this understanding, this paper first overviews the domestic and international development of digital twin cities, summarizes the current representative population baseline mapping methods and their issues, and then proposes an innovative technological framework for mapping real urban population information, aiming to achieve precision in the population baseline. Finally, the paper discusses how this technology can be applied in planning practice.

1. Overview of the Development and Challenges of Digital Twin Cities

1.1 Digital Twin and Digital Twin Cities

A digital twin is a dynamic virtual model of a physical entity created through digital means, capturing multiple dimensions, times, scales, and variables, used to describe, simulate, and mirror the properties, behaviors, and rules of its real-world counterpart. The concept of digital twins was first proposed by Professor Michael Grieves in 2003 during a product lifecycle management course at the University of Michigan, defining it as a digital mirror of physical artifacts or processes. In 2012, NASA proposed one of the most significant definitions of digital twins in its report "Future U.S. Space and Air Force Aircraft Digital Twin Paradigm." After more than a decade of global academic exploration, digital twin theories have flourished. Recently, as China has been vigorously promoting digital transformation, extensive discussions have been held in various fields, with digital twin applications in satellite/space communication networks, ship lifecycle management, vehicle damage resistance, power plant intelligent control, and digital twin cities, among others.

The concept of digital twins entered the urban field in the late 2010s, focusing on virtual-real interaction technologies for urban operation, aiming to achieve spatial-temporal sharing of information throughout the planning, construction, and governance lifecycle, assisting in refined governance decisions and digital resource asset value enhancement. In recent years, developed countries have made considerable progress in urban digital twins, such as Singapore's early "Smart Singapore Plan" (2015), the EU's "Destination Earth" (DestinE) plan (2021), the SmartWorldPro digital twin city platform developed in New Mexico, USA (2019), Japan's Ministry of Land, Infrastructure, Transport, and Tourism's PLATEAU project (2021), and France's government-led Unity platform for urban construction and management (2021). These emerging digital twin projects all focus on addressing public issues, such as: "Destination Earth," which aims to monitor and predict the interaction between human activities and the natural environment to provide decision support for major natural disasters; SmartWorldPro, which focuses on simulating and optimizing urban energy use to promote low-carbon and clean energy adoption; and

PLATEAU, which integrates data from architecture, transportation, geography, economics, health, and disaster prevention to monitor urban activities, respond to disaster events, and simulate planning implementation.

Here is the translation of the provided text into English:

1.2 Challenges of Digital Twin Cities: Urban Complexity and Accurate Mapping of Population Data

The complexity of cities is the biggest challenge in applying digital twin technology to urban areas. Digital twin technology originally emerged in the fields of aerospace and aviation, with its classic applications involving real-time monitoring of aircraft operational states through sensors and computer simulations. However, as Peter Hall pointed out, space programs have clear objectives and resemble fixed targets, making their trajectories predictable. In contrast, urban planning faces the dynamic uncertainty of social factors, making it more akin to targeting moving targets, which is more difficult to implement than space programs. Batty et al. also argued that the mainstream concept of digital twin is an abstraction of a physical system, which is "hard." Mapping a "soft" system, which is driven by human behavior, will be much more difficult. Therefore, the application of digital twin technology in urban planning is fundamentally different from its use in engineering, as it not only needs to map the physical space but also requires a comprehensive focus on the decision-makers and those affected behind the physical space changes, critically responding to the institutional factors of the city, which inherently involve dynamic complexity.

The complexity of cities is also driven by the multiple stakeholders, goals, behaviors, and organizational characteristics of the population, which constitute the "underlying logic" influencing urban development decisions, interactions, and institutional construction. Although some scholars have likened digital twin cities to complex stack models, which are built through vertical, multi-type, multi-scale, multi-attribute, and multi-module operational layers with internal interfaces and protocols for continuous integration and iteration, this does not mean that more data is always better for digital twin cities. Instead, it should map the "required full amount" of data based on actual needs. Among all the city data, population data plays a key role in identifying public issues, analyzing supply-demand relationships, and evaluating urban fairness. It is the most critical data that needs to be fully mapped. Therefore, urban digital twins should first address the authenticity and accuracy of population data mapping to create the conditions for efficient and effective urban public governance. Currently, various methods for mapping urban population data have emerged through exploration in data science, geography, computer vision, and related interdisciplinary fields, providing a foundation for exploring the construction of urban population data.

2. Existing Methods for Mapping Population Data and Their Issues

2.1 Gridding Population Data

Gridding population data into raster units is a common method, which presents the population distribution features in homogeneous units. Existing studies mainly adopt two gridding methods: one is detailed mapping, which involves subdividing census data into grids using auxiliary information such as geographic space data and socio-economic statistics. This method includes techniques such as surface interpolation, statistical regression, machine learning, and stacked integration. This method is relatively mature, with various domestic and international institutions having developed related data products that enable high-resolution grid conversion of census data. The second method is statistical aggregation. With the widespread use of mobile devices, mobile signaling data has also been widely used in population layout research. This method is based on the time and location data left by mobile communication devices, and aggregates the number of individuals in each grid based on regular behavioral patterns to reflect population distribution characteristics such as employment, residence, and recreation.

2.2 Spatialization of Population Data Based on Specific Units

The principles and methods for spatializing population data based on specific units are similar to gridding, but typically involve mapping population data to administrative units (such as communities), transportation zones, medical service areas, school districts, and residential areas. This method helps explore issues related to population distribution and public service performance. For example, studies on urban population spatiotemporal distribution typically use administrative boundaries as mapping units. In transportation service research, transportation zones based on city road networks are used to map population data, which can further explore issues such as work-residence commuting or public transportation service. Research on medical facility layout uses hospital service buffer zones as mapping units to evaluate the accessibility differences in medical services for populations in different areas. Educational facility research uses school districts as mapping units to assess whether population structures and educational demands match the capacity of educational facilities.

In addition to larger spatial units, some studies have refined population data to the scale of building units. For instance, some researchers describe residential spaces using features such as residential block area, building area ratio, number of building floors, and public space ratio, and establish a relationship model between the permanent population and residential space attributes, enabling estimation of the population based on residential block characteristics. Other studies have built relationship models between building volume and static population distribution, and based on this, used spatial geographic data, functional analysis, and vacancy rate identification to estimate population data. These studies refine the spatial unit to residential buildings and use related building indicators as variables to estimate the population. However, the conclusions from these studies represent the "population in the model," not the actual identities of the people.

2.3 Population Data Mapping Based on Urban Sensor Devices

With the development of the Internet of Things (IoT) technology, the infrastructure for urban sensors has been continuously improved, providing new data sources for population data mapping. The IoT architecture consists of three layers: the sensing layer, the internet layer, and the application layer. The sensors located in the sensing layer are mainly responsible for identifying objects and collecting various real-time data, playing a crucial role in population data collection. These data can be categorized into two types: one type is image and video data collected by surveillance equipment, which can estimate the population size through computer vision analysis. In urban planning and public service fields, researchers further perform spatial mapping using the geographic information of image data, storing it as GIS data and aggregating it in defined spatial units. The second type is data from non-visual sensor facilities, such as Wi-Fi signals, radio frequency signals, and environmental data, which are rich in types and relatively high in accuracy. These can also be used for mapping population data. For example, Wi-Fi probes estimate the population by tracking the number of visits from mobile devices based on Wi-Fi signals, while exit radio frequency devices estimate the population by counting the frequency of signal interruptions. Environmental interaction devices estimate the population by collecting data on interactions between people and the built environment.

2.4 Summary of Existing Problems: Inaccuracy, Lack of Authenticity, and Insufficient Coverage in Population Base Mapping

The three methods for mapping population data mentioned above have been widely used in research and practice. However, from the perspective of establishing the "population base" needed for urban operation and management, current methods still face issues in terms of accuracy, authenticity, and coverage.

Firstly, the grid-based method abstracts population data into grid units, which has value at a macro scale, but suffers from distortion at the meso and micro scales. This is because the method overlooks the specific impact that the complex built environment has on population distribution.

Secondly, the method based on specific units (e.g., administrative districts, traffic zones, etc.) takes the complexity of the built environment and its morphological features into account, and the computational accuracy is improving. However, this method typically uses statistical models to estimate population numbers, neglecting the local identity of the population. This leads to an inability to accurately assess residents' public service needs.

Thirdly, methods that use sensor devices to detect crowds are mainly applied in key urban areas, which are advantageous for urban design, public safety monitoring, and similar tasks. However, as the detection range is often limited to specific regions, these methods are not suitable for meeting the "full coverage" base data requirements of urban planning, construction, and governance.

Therefore, there is an urgent need for a method that balances accuracy, authenticity, and full coverage to meet the demands of urban life-cycle operation and management in the digital twin era.

3. Innovative Technical Framework for Accurately Mapping Urban Population Base

With the digital transformation of urban governments in recent years, a large amount of citizen service data has been stored on government digital platforms, leading to the emergence of "government data" as a new source of information. This type of data provides a foundation for achieving realistic, fine-grained, and complete population mapping. In recent years, my research team has carried out urban digital twin practices in various districts of Pudong, Huangpu, and Jing'an. Through continuous validation, adjustment, and optimization, we have developed an operational and innovative technical framework for constructing urban population bases. In this process, we have established principles based on "realism" and extended the logic of "space-population-demand" into three parts: "mapping real space," "depicting real population," and "responding to real needs" (see Figure 1).

(1) Mapping Real Space: Construct a "fine-grained" spatial base map for the city's physical environment. This involves two main components: (1) geographical data for different administrative boundaries, along with built environment features such as blocks, parcels, and buildings; and (2) household-level address data based on building units, which is obtained through various channels of government data. By aggregating administrative boundary data, built environment data, and household address data, we can create a "precise-to-household" fine-grained spatial base map. This approach not only avoids the distortion issues of grid-based data but also improves the data's granularity and authenticity.

(2) Depicting Real Population: Use multi-line governmental data to obtain actual population information and link it to the spatial base map to achieve "precise-to-household" population data. This is followed by integrating multi-source population big data to enrich and supplement the population information on the government population base, and then constructing a life-cycle population profile to accurately depict the population. This process ensures that the real identity information of the population is appropriately portrayed, while respecting legal privacy.

(3) Responding to Real Needs: Use a graph database as the engineering solution for the "space-population" base, forming a data platform product, along with supporting technologies and mechanisms for data maintenance, monitoring, and updating. Based on this foundation, specific datasets can be extracted from the graph database for different types of service scenarios (e.g., livelihood, education, healthcare, community management) to support applications in various public services.

3.1 Mapping Real Space: Constructing a "Precise-to-Household" Spatial Base Map

3.1.1 Constructing a Spatial Base Map Based on Built Environment and Administrative Boundaries

For any non-outdoor activities undertaken by the urban population, they must rely on the built environment (e.g., residential, office, and commercial functions) as a carrier. Simultaneously, from the perspective of governance, there exists a "city" composed of an administrative boundary system. This system is used by various government levels to manage public affairs within their jurisdiction. Therefore, the construction of a spatial base map should include both systems: on one hand, for the built environment, this includes spatial models of individual buildings, blocks, streets, functional areas, city districts, and even the entire city; on the other hand, for administrative boundaries, this includes floor levels, community management "micro-grids," administrative districts, towns, neighborhoods, city borders, and provincial boundaries, which correspond to administrative governance units and dynamically adjust with changes in administrative areas.

The technical path for mapping the spatial base map is as follows:

1. Collect and organize relevant data from various map resources. Administrative boundary data primarily comes from government public information, while physical environment data is sourced from geographic data from surveying agencies, offline records in government work, and internet maps.

2. Standardize the collected data into a unified format. If administrative boundary data is missing, it can be supplemented or corrected manually; for non-government internet data with diverse formats, AI, image recognition, and information extraction technologies can be used for processing. Regular updates from surveying agencies and validation from governmental workers ensure that the spatial base map remains accurate.

3. Spatially associate the standardized geographic data according to the actual

situation, ensuring the correct hierarchical relationships. On one hand, each administrative level should seamlessly cover all subunits beneath it; on the other hand, building models should cover all floor and household information. See Figure 2.

3.1.2 Refining the "Household" Information in the Spatial Base Map

An accurate spatial base map must rely on real address information for support, which is typically represented by the "household" unit within a building. Although the "household" serves as the spatial carrier for the resident population, such data generally does not appear in surveying or mapping data and can only be obtained through other channels. Based on current practical experience, obtaining "household" data in the spatial base map is typically divided into two situations: The first situation is that some large cities have established specialized building and housing basic data in government systems, such as housing databases retained by housing management departments, self-constructed housing databases by street or neighborhood committees (in the form of databases or paper-based housing status maps), and "community clouds" (Shanghai) or "four grassroots governance platforms" (Hangzhou) systems, which include accurate location and address data, continuously updated through grassroots visits. In this case, data can be directly obtained and integrated into the spatial base map. The second situation is that some regions lack a complete building and housing database, requiring the use of "population household address technology" to acquire household data. The principle is to reverse-engineer the "household" data by collecting resident address information from government population data. Specifically, the address data is first collected, then parsed using natural language processing techniques like dictionary-based word segmentation, recognizing the address structure ("Province-City-District-Road-Alley-Number-Room") to extract the precise "household" information, which is then linked to the building household model and integrated into the spatial base map.

3.2 Depicting Real Populations: Population Attachment to Space and Establishing Portrait Systems

Accurate population information is the foundation for analyzing public service issues, and it includes two key aspects: First, the "local" information of the population, achieved by linking it with the spatial base map. Second, the depiction of population characteristics, which involves how to create a population portrait. This paper proposes the idea of a "lifecycle population portrait map" and, by integrating multi-source non-governmental data, further supplements the "thematic portraits." The steps above collectively form the framework for mapping a refined population base. See Figure 3.

3.2.1 Population Data Attachment to the Spatial Base Map

Population data mainly comes from various governmental data sources, such as household registration data from the police department, birth and health record data from the health commission, marriage and pension data from the civil affairs bureau, and employment and social security data from the human resources department. At the same time, companies, as social entities, leave corresponding corporate data when registering businesses, applying for patents, or engaging in investment and financing activities. Furthermore, data departments or urban management departments have also created various digital application scenarios, accumulating population activity data such as work order hotlines and transaction consultations. Building a refined population base requires integrating data from these various government departments. During integration, different data sources are assigned different levels of priority to avoid duplicate data collection. After the integration of government data, the population data is attached to the spatial base map using the address field. For example, household information from the population's residence address is linked to the corresponding building's spatial data; similarly, enterprise credit codes from population social security data are linked to corresponding office building spatial data.

3.2.2 Full Lifecycle Population Portraits

Identity information is key to reflecting the "authenticity" of population data, and necessary identity information also meets the needs of efficient government management and precise public services. Based on the "full lifecycle management" concept, a comprehensive "population portrait map" framework has been established, mainly referencing the following three aspects: First, academic research on population lifecycle, family lifecycle, and population portraits; Second, current standards for population coding and classification, including national, industry, and local standards; Third, the current state of data classification in administrative departments, as well as data application scenarios based on departmental business needs. This research draws mainly on the team's investigations of population database construction in cities such as Shanghai, Beijing, and Hangzhou. Based on this research, a "1+9+X" population portrait map system is proposed: "1" refers to the basic attributes of the population, including gender, age, ethnicity, and local household status, where "local household status" indicates whether a resident "lives locally" based on community records or whether they are "registered locally" according to the police household registration. "9" refers to the multiple population information dimensions closely related to governance scenarios under the "full lifecycle management" perspective, including education, employment, marriage and childbearing, pension, social activities, and social identity. "X" refers to population attributes under specific or special conditions, which could include public events with sudden time-sensitive attributes, such as a person's "positive/negative test result" in a public health emergency, or population attributes closely related to regional traditions, such as being an "important elder" in an ethnic community. The "X" in the portrait map system is open-ended and can be flexibly adjusted based on actual situations.

3.2.3 Multi-Source Data Integration

Government data can generally describe the city's permanent and working population, but there is a lack of description for populations engaged in other diverse activities in the city, such as leisure, tourism, and consumption. To portray these populations, additional non-governmental commercial big data is needed. For example, analyzing mobile signaling data with modeling algorithms can depict the distribution of leisure populations in the city, showing resident, transient, and mobile population activity patterns. Similarly, data from internet platforms related to consumption, reviews, and social media can describe the activity levels and spatiotemporal distribution of the city's consumer population. By comparing and integrating these multi-source big data from non-governmental sources with governmental population base data, it is possible to supplement and improve the population data dimensions, as well as to create "thematic" portraits of the population.

3.3 Responding to Real Demands: Achieving Multi-Scenario Government Services through Data Engineering Methods

Different departments and public institutions within urban governments have varying demands for population data due to specific service scenarios. To meet the needs of different departments with "multiple users, multiple scenarios, and high variability," the research team has developed an engineering method that processes population data into a digital platform product capable of supporting daily use for different users and enabling a fully automated loop for data "integration—governance—maintenance—update." This engineering method includes the "Database Architecture System" and the "Data Operation and Maintenance System" (Figure 4). The former supports the data correlation network and mapping process of the population data, as well as data application requests from different user ends; the latter ensures the accuracy, authenticity, and timeliness of the population data through a technical mechanism.

The "Database Architecture System" is an engineering method for synchronously updating and flexibly expanding population data alongside complex urban entities. Therefore, this database architecture needs to be highly universal, minimally constrained, and highly efficient in integration. Since urban population data involves complex correlations and dynamic changes between "space—population—demand," using traditional relational databases could lead to issues such as slow data retrieval, poor product performance, and users being unable to access the system during modifications. As a result, a graph database with an "entity—relation" structure is chosen as the underlying architecture of the product. In the detailed population data, physical spaces like buildings, and social entities such as populations and businesses, are treated as "entities," and the relationships between them are considered "relations." For example, a citizen "lives in" a building, "works in" a company, or "is registered in" a specific organization, forming associations between entities. When any data changes, only the attributes of the "relations" need to be modified. This simple "entity—relation" data structure of the graph database can support the expansion and flexible application of the population data platform for various user endpoints.

The “Data Operation and Maintenance System” consists of three parts: data governance engineering, data quality monitoring, and data maintenance closed-loop. Data governance engineering is the “production flow” of the data baseline, after which the data enters the graph database. Through steps such as initial data entry, data cleaning and correction, aggregation and statistical calculation, population profiling and tagging, and data splitting for application, the population baseline can be directly applied to various government service scenarios when the data is extracted from the system. Data quality monitoring is the “monitoring flow” of the data, ensuring the consistency, accuracy, and timeliness of data through automated and manual methods, as well as monitoring business applications. This ensures the quality of the extracted data. The data maintenance closed-loop is the “update flow” of the data. The closed-loop operates in two directions: the upward loop involves regular updates and issue feedback between the data and higher-level data sources (i.e., data centers or business departments); the downward loop refers to updates, verification, and corrections made by users during frontline usage.

4 Outlook on the Application of Detailed Population Data

4.1 Necessity of Application and Data Security Assurance

The mapping process of detailed population data involves a large amount of private information, such as age, identity, address, health, and whereabouts. According to the Civil Code, personal data such as names, birth dates, IDs, biometric data, addresses, phone numbers, emails, health information, and movement data are legally protected. Therefore, two issues arise: 1) Given the vast amount of private data involved, is its use necessary? 2) If necessary, how can data security be ensured?

Regarding the first issue, it is found that Chinese law encourages the use of personal data for public services and public affairs. For example, the Civil Code specifies that the collection of private data to maintain public interests and protect personal rights does not require civil liability. The Data Security Law further clarifies that “the state supports the development and utilization of data resources to enhance the intelligence of public services, taking into full account the needs of the elderly and disabled.” From the perspective of optimizing the quality of urban public services and improving urban governance, mapping the “full population” is necessary.

For the second issue, relevant practical experiences provide solutions: Detailed population data includes two stages: “development” and “application.” The “development” stage involves mapping multi-source population information into a data platform that operates in a government network environment and does not involve “users,” so data security can be ensured by isolating the network environment. The “application” stage involves a large number of “users” interacting with the system, and since the operating environment is relatively open, data security can be ensured by introducing strict access control systems, management standards, and accountability mechanisms.

4.2 Application of Detailed Population Data in Urban Population Analysis

Detailed population data comes from numerous government and non-government data sources, and the update frequencies of these data sources vary: 1) Government data has an "uncertain" update frequency. For example, data regarding household registration, children's enrollment, unemployment benefits, outpatient treatment, and social security usage are updated in corresponding government systems, with update frequencies directly related to the occurrence of specific events. Household registration and school enrollment are typically low-frequency events, while unemployment benefits and medical/social security data are relatively high-frequency. 2) The frequency of grassroots work logs depends on the frequency of visits by local workers, and different populations have different update frequencies, with more frequent visits and data updates for key groups. 3) Non-government data, including mobile signaling, GPS movements, and social media, typically records data during mobile terminal use. When the platform provides data, it aggregates it into hourly, daily, weekly, or monthly intervals as needed. Therefore, the update frequencies of different types of data sources may vary depending on events, individuals, and requirements.

Integrating the "multi-source and variable-frequency" demographic data into a refined population database platform

The integration of multi-source and variable-frequency demographic data into the refined population database platform allows for two primary benefits:

1. **Spatial Aggregation for Frequency Analysis:** The data can be aggregated into spatial units of varying scales to analyze frequency differences in specific demographic data across different spatial units.

2. **Multidimensional Spatiotemporal Population Analysis:** From different perspectives, the platform enables the analysis of the spatiotemporal characteristics of urban populations, including:

- **Based on administrative data:** It is possible to analyze the spatiotemporal features of residential registration, spatial flows of children entering kindergartens and schools, the spatiotemporal differentiation of employed and unemployed residents, and the temporal patterns of residents' use of medical facilities. This provides evidence for developing urban public policies.

- **Based on grassroots work records:** The platform allows for precise statistics on residents' registered addresses and places of residence, identifying categories such as "residents living at registered addresses," "residents living away from registered addresses," and "residents absent but still registered." Among these:

- “Residents living at registered addresses” are part of the stationary population.
- “Residents living away from registered addresses” can be categorized into permanent or mobile populations depending on whether their length of residence exceeds six months.
- “Residents absent but still registered” are classified as the outflow population.

This enables fine-grained data to support analyses of the permanent population and actual residents.

- **Based on non-governmental data:** This type of data has already been widely applied in urban density analysis, vitality assessment, and structure identification. Incorporating non-governmental data as supplementary information into the refined population database framework expands the informational dimensions of urban demographic mapping.

4.3 Application of the Refined Population Database in Community Governance Planning

Communities are the fundamental units of spatial governance, and precise public service delivery is one of the primary tasks of community governance. Against the backdrop of the complex and dynamic “space-population” urban environment, precise provision of public goods in communities is a key factor in improving residents’ quality of life and ensuring the efficient use of limited public resources. The refined population database provides a scientific basis for this purpose. Decision-makers can assess public service needs based on the characteristics of different service groups and formulate corresponding public goods supply plans in the following areas:

1. **Facilities and Spatial Public Goods Supply:**

By combining refined population data with city road networks, public facility POI data, and incorporating algorithms and models, the platform supports various scenarios for providing facilities and spatial public goods. For example:

- **15-Minute Community Life Circle Planning and Monitoring:** Using spatial algorithms to calculate actual walking path coverage and combining refined population data, the platform can accurately estimate the scale of subpopulation groups served by facilities and determine facility sizes based on their specific needs.
- **Educational Resource Planning:** Combining population data with educational statistics, the platform can model scenarios such as school admissions, predict school-age populations, and evaluate resource allocation for education. Data can be shared with citizens via mini-programs, aiding decisions related to real estate purchases and children’s education.

2. **Soft Service Provision for Targeted Groups:**

Population profiling tags in the refined population database facilitate precise targeting of specific service groups, enabling customized service delivery. For instance:

- **Assistance for Vulnerable Groups:** Community staff can identify individuals meeting aid criteria, such as elderly residents, persons with disabilities, or unemployed citizens, and match them with social management resources for precise assistance delivery.

- **Healthcare Services:** For enhancing the reach of community hospital services, the platform can identify key populations such as pregnant women or critically ill patients, compare them with hospital records, and encourage service registration or offer tailored health consultations and reminders.

3. **Public Participation and Other Service-Oriented Public Goods:**

Integrating refined population data with online interactive formats facilitates emerging democratic participation models. Traditional offline participation mechanisms often limit engagement due to outdated systems. Leveraging refined population data and technologies like virtual reality, digital avatars, and real-time rendering, specific groups can securely participate in residents' meetings or homeowners' assemblies online, breaking time-space constraints and increasing civic participation.

4.4 Application of the Refined Population Database in Territorial Spatial Planning

As the territorial spatial planning system matures, higher standards are set for urban digital twin initiatives. The refined population database, accurately mapping the real "space-population" system and aggregating household-level data into spatial governance units of varying scales, provides precise foundational data for various planning efforts:

1. **Comprehensive Planning:**

For holistic management of all elements within territorial space, the database offers a complete population dataset and profiling maps. These support tasks like structural relationship identification, functional layout optimization, and dynamic plan adjustments by providing accurate and timely information on population distribution, changes, and flows, enhancing the efficacy of spatial planning decisions.

2. **Detailed Planning:**

With the trend toward "unit planning + implementation plans" in detailed spatial planning, the high-resolution "space-population" foundation provided by the database supports standard-setting, boundary selection, and adjustment for planning units. It also aids in identifying issues, recognizing gaps, and formulating implementation plans, promoting refined management and digital governance of planning units.

3. **Diagnostics and Evaluation:**

Population information is often inadequately integrated with built environments, operational systems, and activity systems. Using household-level data aggregated into grids, blocks, districts, or city zones, the refined database accurately maps multidimensional population information within specific areas. This enables analysis of "space-population" compatibility, providing insights for optimizing systems such as ecological environments, housing security, public services, and transportation infrastructure, ensuring a precise match between population activities and supporting systems.

4) Implementation Monitoring

With the introduction of the National Territorial Space Planning Implementation Monitoring Network (CSPON), the establishment of a "single map" information platform at the national, provincial, municipal, and county levels using digital methods has become an important task. This platform collaborates with multiple departments and horizontally connects various types of business systems, creating a new open governance ecosystem based on co-construction, co-management, and shared governancened population base can serve as one of the business systems horizontally connected within the CSPON platform. It not only provides accurate population base data to the CSPON platform but also overlaps and integrates with other data within the platform, enabling precise tracking of the "space-population" interaction relationship and enhancing the effectiveness of implementation monitoring.

5. Conclusion

The digital twin of a city does not need to replicate every element in the real city. As Batty et al. [4] pointed out, creating a twin that exactly mirrors the real-world system has no meaning; otherwise, the twin itself would become the real system. The concept of a digital twin should involve the integration of the twin with the real system in some way, but it should not become the real system itself. In this sense, the "digital twin" of a city should be used as a "foundation data mapping integration" for urban planning, construction, and governance. It provides the scientific basis for decision-makers to analyze public issues and develop improvement policies. Among the many objects that can be digitized, the accurate population base of a city is the most critical and challenging part. Based on an analysis of existing methods and their issues, this paper, through practical exploration, constructs a refined population base mapping path that integrates map data, government data, profile labels, and commercial big data, based on the principles of "real space – real population – real demand." It also discusses how this technology can empower future urban planning practices. This framework is proposed in response to the urgent need for digital transformation in urban planning, construction, and governance, reflecting the original mission of "people-centered" urban and rural planning, and contributing to the implementation of the new development concept of "people's cities."

Thanks to the technical support of the team at Shanghai Maice Data Technology Co., Ltd., and the assistance from leaders in digital city projects in Pudong, Jing'an, and Huangpu; and thanks to the reviewers for their insightful comments and valuable suggestions.

Footnotes

1. NASA's definition: A digital twin is a multi-physical, multi-scale, probabilistic integrated simulation of an as-built vehicle or system, using existing best physical models, new sensors, historical data, etc., to reflect the full lifecycle usage status of the twin object. Since this milestone definition was introduced, various definitions of digital twins have begun to emerge in different fields.

2. A Graph Database (GDB) is a concept in computer science. It is a database that uses a "graph structure" for semantic queries, representing and storing data using nodes, edges, and attributes. The key concept of this system is the graph, which directly links data items in storage with nodes and the set of edges representing the relationships between the nodes. These relationships allow data to be directly linked together in the storage area and retrieved in many cases with a single operation. Graph databases prioritize relationships between data.

3. "Microgrid Actual Grid" refers to maintaining the existing grid working level and further dividing the governance scope, focusing on buildings, units, and forming three-level divisions of the community's overall grid, sub-district grid, and building microgrid, and exclusive grids for self-division. This aims to achieve full coverage and refinement of grassroots governance. More information is available in the article from "China Youth Online."

4. Offline ledgers refer to non-networked government work ledgers. These ledgers are usually used to record various tasks, project progress, meeting minutes, decision implementation, work summaries, etc. They can be in traditional paper form or electronic spreadsheets, currently mainly in offline Excel formats.

5. Based on current experience, data from the Housing Management Bureau's building housing database is difficult to obtain and is rarely used in practice.

6. Dictionary-based word segmentation is a common segmentation algorithm in computer natural language processing (NLP). Its principle is to use an existing dictionary corpus to identify the text that matches dictionary entries and label them individually. This method allows for quick segmentation of text but is limited by the scope of the dictionary corpus.

7. Relational Databases (RDB) are a method of building information structures in tables, rows, and columns. RDBs can establish relationships or associations between information by joining tables, enabling users to understand and analyze the relationships between various data points. The drawback of relational databases is that they perform poorly when handling high-concurrency reads and writes, variable fields, or large data writes.

References

[1] Zhuang Shaoqin, Zhao Xingshuo, Li Chenyuan. The Dimensions and Temperature of

National Territorial Space Planning. Urban Planning, 2020, 44(1): 9-13.

[2] Dang Anrong, Tian Ying, Li Juan, et al. The Development Process and Outlook of Smart Territorial Space Planning Management in China. Science and Technology Review, 2022, 40(13): 75-85.

[3] Tao Fei, Liu Weiran, Zhang Meng, et al. The Five-Dimensional Model of Digital Twins and Applications in Ten Fields. Computer Integrated Manufacturing Systems, 2019, 25(1): 1-18.

[4] Michael Batty, Lin Xuhui. Digital Twins, Turing Tests, and Urban Models. Shanghai Urban Planning, 2023(5): 1-3.

[5] FULLER A, FAN Z, DAY C, et al. Digital Twin: Enabling Technologies, Challenges, and Open Research. IEEE Access, 2020, 8: 108952-108971.

[6] SEMERARO C, LEZOCHE M, PANETTO H, et al. Digital Twin Paradigm: A Systematic Literature Review. Computers in Industry, 2021, 130: 103469.

[7] Wan Li, Yin Luoyi, Tang Junqing, et al. Critical Thinking on the Application of Digital Twins in Urban Planning Practice. Shanghai Urban Planning, 2023(5): 18-23.

[8] Yang Tao, Tian Ying, Xu Yanjie. Interactive Generative Planning and Governance Enabled by Digital Twins. Shanghai Urban Planning, 2023(5): 4-10.

[9] Yi Xueqin. Domestic and International Experiences and Insights on Digital Twin Cities. Information and Communication Technology and Policy, 2023, 49(8): 25-30.

[10] Wu Zhiqiang, Gan Wei, Zang Wei, et al. The Concept and Development of City Intelligent Models (CIM). Urban Planning, 2021, 45(4): 106-113.

[11] Yang Tao, Yang Baojun, Bao Qiaoling, et al. Digital Twin Cities and City Information Models (CIM): A Case Study of the Xiong'an New Area BIM Management Platform Project. Urban and Rural Construction, 2021(2): 34-37.

[12] Zhou Yu, Liu Chuncheng. The Logic and Innovation of Building a Digital Twin City in Xiong'an New Area. Urban Development Studies, 2018, 25(10): 60-67. [13] Dang Anrong, Wang Feifei, Qu Wei, et al. Urban Information Modeling (CIM) Empowering the Development of New Smart Cities: A Review [J]. Chinese Famous Cities, 2022, 36(1): 40-45.

[14] Yang Junyan. From Digital Design to Digital Management: The Exploration of the Fourth Generation Urban Design Model in Weihai [J]. Urban Planning Journal, 2020(2): 109-118.

[15] Yang Baojun, Yang Tao, Feng Zhenhua, et al. Digital Planning Platform: A New Model for Serving Future Urban Planning and Design [J]. Urban Planning, 2022, 46(9): 7-12.

[16] Zheng Degao, Lin Chenhui, Wu Hao, et al. Spatial Research and Digital Portrait Technology Framework for Sustainable Urban Development [J]. Urban Planning Journal, 2023(6): 32-39.

[17] Yang Tao, Li Jing, Li Mengyao, et al. Digital Twin Methods for the Protection and Revitalization of Suzhou's Ancient City Historical and Cultural Heritage [J]. Urban Planning Journal, 2024(1): 82-90.

[18] Wu Zhiqiang, Zhou Mimi, Liu Qi, et al. "Intergenerational Twin": Reflecting the Life Characteristics of Cities [J]. Urban Planning Journal, 2024(1): 9-17.

[19] Tian Ying, Yang Tao, Dang Anrong. The Logic of Digital Twin City Construction Based on Scene Iteration [J]. Shanghai Urban Planning, 2023(5): 24-30.

[20] Han Tao, Guo Xi. From Cultural Twin to Technological Twin to Digital Twin: A Historical Overview of Digital Twin City Logic Based on Big History [J]. Shanghai Urban Planning, 2023(5): 31-35.

[21] Wang Xuemei, Li Xin, Ma Mingguo. Progress and Case Studies on Population Data Spatialization Based on Remote Sensing and GIS [J]. Remote Sensing Technology and Applications, 2004(5): 320-327.

[22] Cheng L, Wang L, Feng R, et al. Remote Sensing and Social Sensing Data Fusion for Fine-Resolution Population Mapping with a Multimodel Neural Network [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 5973-5987.

[23] Niu Xinyi, Lin Shijia. Spatiotemporal Big Data in Urban Planning Research: Technological Evolution, Research Topics, and Emerging Trends [J]. Urban Planning Journal, 2022(6): 50-57.

[24] Ding Liang, Niu Xinyi, Song Xiaodong. Progress in Urban Space Research Based on Mobile Location Big Data [J]. International Urban Planning, 2015, 30(4): 53-58.

[25] Li Xingyue, Chen Fulin. Study on Urban Connections and Population Research in Small and Medium-sized Urban Areas Based on Mobile Phone Signaling Data [J]. Urban Transportation, 2020, 18(4): 47-54.

[26] Li Fengqing, Zhao Min, Wu Mengdi, et al. Spatial Performance Mechanism of the "Multi-center" Spatial Structure of Large Cities: A Study Based on Xiamen LBS Data and Regular Census Data [J]. Urban Planning Journal, 2017(5): 21-32.

[27] Wang De, Liu Zhenyu, Yu Xiaotian, et al. Strategic Perspective Analysis of Urban Population Size: A Case Study of Wuhan Population Size [J]. Urban Planning Journal, 2017(5): 58-65.

[28] Long Ying, Zhang Yu, Cui Chengyin. Analysis of Beijing's Job-Housing Relationship and Commuting Travel Using Bus Card Data [J]. Geographical Journal, 2012, 67(10): 1339-1352.

[29] Teh B T, Shinozaki M, Chau L W, et al. Using Building Floor Space for Station Area Population and Employment Estimation [J]. Urban Science, 2019, 3(1): 12.

[30] Feng Mingxiang, Fang Zhiyang, Lu Xiongbo, et al. COVID-19 Spatiotemporal Diffusion Estimation Method at Traffic Analysis Zone Scale: A Case Study of Wuhan [J]. Journal of Wuhan University (Information Science Edition), 2020, 45(5): 651-657.

[31] Liu Xiaocong, Liu Yongwei, Cai Fei, et al. City Emergency Medical Facility Recommendation Method Based on Spatiotemporal Accessibility [J]. Journal of Earth Information Science, 2019, 21(9): 1411-1419.

[32] Wesołowski A, O'Meara W P, Tatem A J, et al. Quantifying the Impact of Accessibility on Preventive Healthcare in Sub-Saharan Africa Using Mobile Phone Data [J]. Epidemiology (Cambridge, Mass.), 2015, 26(2): 223.

[33] Xu Tao, Liu Yingying, Lu Guiyi, et al. Urban Education Facility Configuration Method Based on Population Spatial Differentiation: A Case Study of Qiaokou District Junior High Schools in Wuhan [J]. Modern Urban Research, 2022, 7(37): 93-99.

[34] Dong Nan, Yang Xiaohuan, Cai Hongyan. Study on Population Data Spatialization Method Based on Residential Space Attributes [J]. Geographical Science Progress, 2016, 35(11): 1317-1328.

[35] Qiu F, Sridharan H, Chun Y. Spatial Autoregressive Model for Population Estimation at the Census Block Level Using LIDAR-Derived Building Volume Information [J]. Cartography and Geographic Information Science, 2010, 37(3): 239-257.

[36] Urals, Hussain E, Shan J. Building Population Mapping with Aerial Imagery and GIS Data [J]. International Journal of Applied Earth Observation and Geoinformation, 2011, 13(6): 841-852.

[37] Bakillah M, Liang S, Mobasheri A, et al. Fine-resolution Population Mapping Using OpenStreetMap Points-of-Interest [J]. International Journal of Geographical Information Science, 2014, 28(9): 1940-1963.

[38] Lian Ting. Population Estimation at the Building Scale Based on Random Forest and Nighttime Light Data [D]. Shanghai: East China Normal University, 2019.

[39] Shang S, Du S, Du S, et al. Estimating Building-Scale Population Using Multi-Source Spatial Data [J]. Cities, 2021, 111: 103002.

[40] Garrido-Valenzuela F, Cats O, Van Cranenburgh S. Where Are the People? Counting People in Millions of Street-Level Images to Explore Associations Between People's Urban Density and Urban Characteristics [J]. Computers, Environment, and Urban Systems, 2023, 102: 101971.

[41] Long Ying, Cao Zhejing. Self-feedback City Design Method Based on Sensor Devices and Online Platforms and Its Practice [J]. International Urban Planning, 2018, 33(1): 34-42.

[42] Li Li, Zhang Jing, Fang Lixin. Study on Low-Precision Wi-Fi Probe Data Collection and Analysis Method: A Case Study of Environmental Behavior at Block Scale [C]// Smart Future: Proceedings of the 2021 National Conference on Architectural Digital Technology Teaching and Research, 2021.

[43] Li Ruping, He Ziqi, Zhang Linyan, et al. Crowd Flow Monitoring System Based on WeChat Mini Program Cloud Development [J]. Electronic Technology and Software Engineering, 2021(8): 68-70.

[44] Kannan P G, Venkatagiri S P, Chan M C, et al. Low-Cost Crowd Counting Using Audio Tones [C]// Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems, 2012: 155-168.

[45] Zhang Junjun, Shi Zhiguang, Li Jicheng. Current Status and Trends of Population Counting and Crowd Density Estimation Techniques [J]. Computer Engineering and Science, 2018, 40(2): 282-291.

[46] Li Jianing, Zhang Yiping, Tang Ge, et al. Initial Construction of a Digital Twin City Working System for Grassroots Governance: A Case Study of Huamu Digital Twin City [J]. Shanghai Urban Planning, 2023(6): 91-97.

[47] Chen Si. Spatial Comparison Analysis Model Based on the Population Lifecycle [J]. Geospatial Information, 2020, 18(12): 24-26.

[48] National People's Congress of the People's Republic of China. Civil Code of the People's Republic of China [S]. 2020-05-28.

[49] National People's Congress of the People's Republic of China. Data Security Law of the People's Republic of China [S]. 2021-06-10.

[50] Wang De, Ren Xiyuan. Distribution and Mobility of Shanghai's Permanent Population from the Perspective of Daily Flow [J]. Urban Planning Journal, 2019(2): 36-43.

[51] Zhang Shangwu, Yan Longxu, Wang De, et al. Policy Path Analysis for Spatial Structure Optimization of the Shanghai Metropolitan Area: A Scenario-Based Analysis of Population Distribution [J]. Urban Planning Journal

[52] Zhang Shangwu, Liu Zhenyu, Zhang Hao. Discussion on Detailed Planning and Its Operational Model under the National Spatial Planning System [J]. Urban Planning Journal, 2023(4): 12-17.

[53] Wu Jiang, Wang Xin, Chen Ye, et al. The Challenges of Urban Check-ups for Megacities and Shanghai's Practice [J]. Urban Planning Journal, 2022(4): 28-34.

[54] Wang Wei, Liu Ze, Lin Yuxian, et al. From "One Map" of National Spatial Planning to CSPON "One Network": An Academic Dialogue [J/OL]. Beijing Planning and Construction, 2024-01-17: 1-39. http://kns.cnki.net/kcms/detail/11.2882.tu.20240110.1523.002.html.

[55] Wang Wei, Liu Ze, Lin Yuxian, et al. From the "One Map" of National Spatial Planning to the "One Network" of CSPON: An Academic Dialogue [J/OL]. Beijing Planning and Construction, 2024-01-17: 1-39.
http://kns.cnki.net/kcms/detail/11.2882.tu.20240110.1523.002.html.