Assisting refined urban management: Building an evaluation framework of data mapping rate towards digital twin city platform

Xinghan Chen[1] **Yiping Zhang**[2] **Yu Ye**[3]**.**

Abstract. In recent years, digital twin city platforms often encounter issues such as emphasizing the physical model's accuracy over social cognition and specialized applications over a comprehensive data system, hindering the fulfilment of refined urban management's real needs. Therefore, it is essential to define the characteristics of an urban management-oriented digital twin platform and construct a detailed evaluation mechanism. This study examines the framework for evaluating mapping rates, introducing three indicators: data resolution, data freshness, and data relevance. We developed a quantifiable and replicable evaluation model to assess data completeness, update timeliness, and network correlation degree. Using Shanghai's Huamu digital twin platform as a case study, we calculated each indicator and formed a comprehensive mapping rate evaluation. This research achieves a quantitative analysis of digital twin city platforms' development quality which was previously unmeasurable. Additionally, this study aids in advancing digital twin city platforms to facilitate the development of a "bottom-up" refined urban management approach.

Keywords: digital twin city; refined urban management; data evaluation model; digital twin mapping rate.

---

[1] Xinghan Chen
Department of Architecture, College of Architecture and Urban Planning, Tongji University, Shanghai, China
e-mail: 2230078@tongji.edu.cn

[2] Yiping Zhang
MetroDataTech Ltd., Shanghai, China
e-mail: zhangyiping@metrodata.cn

[3] Yu Ye (✉)
Department of Architecture, College of Architecture and Urban Planning, Tongji University, Shanghai, China
e-mail: yye@tongji.edu.cn

1 Background

In the context of the system reform of refined urban management (3), advances in information technology and big data offer new prospects for evolving management models (2, 7). The digital twin city concept represents a cutting-edge urban management paradigm, merging diverse urban data to synergize digital and physical city spaces (9, 18). This approach aims to enhance urban management, particularly in strategic decision-making for urban challenges like pollution, congestion, energy, and land use (11, 14). In the early stages, digital twin city projects focused on data collection and visualization, utilizing municipal cloud services and technologies like 3D modelling and real-time rendering to create visual representations for urban planning and management (5, 15).

Despite progress in developing digital twin cities, current endeavors are in their infancy and encounter two primary challenges. The first is an "overemphasis on physical precision at the expense of socio-economic digitization." Present digital twin cities prioritize replicating the physical aspects of urban environments, neglecting the complex socioeconomic activities within the city (19). This approach often does not align with the practical needs of market entities and government regulators (13). Moreover, the collection, purification, and integration of socioeconomic data are intrinsically more intricate, necessitating a comprehensive and methodical understanding. The second issue, "favoring scenario-specific applications over integrated data systems," stems from a tendency in digital twin city platform construction to focus on narrow use cases, leading to fragmentation and isolation, further impeding holistic urban management (17).

These issues suggest that the construction of digital twin platforms has neglected the degree of correspondence between the physical and virtual representations of cities. As a response, our study introduces the concept of a "mapping rate" for digital twin platforms, accompanied by a systematic methodology. This framework examines critical aspects such as urban space, infrastructure, socioeconomic factors, and information dynamics, assessing their alignment and integration with their digital counterparts. Our objective is to gauge the maturity of digital twin platforms and to establish a suite of assessment tools that are practical and scalable.

2 Data Mapping Rate

Facing the demands of refined urban management, it's crucial to precisely define and scrutinize the concept of data mapping rate and identify relevant indicators (see Figure 1). First of all, based on the nature of the digital twin, the assessment should concentrate on the detail and depth with which digital twin platforms replicate spatial, social, and economic information (1). Accordingly, this study presents the notion of 'data resolution' to evaluate the comprehensiveness of data collection. Additionally, given the dynamic nature of data within urban management (6), the platform must ensure timely updates to fulfil the stringent demands for accuracy and efficiency in sophisticated urban management. This leads to the proposition of a 'data freshness' indicator, designed to evaluate if data is current or outdated, emphasizing the alignment of update frequencies with real-world needs. To transcend siloed approaches, data across various sectors must be integrated as per operational necessities, establishing a cohesive management network (10). Thus, this study introduces the concept of 'data relevance,' aimed at assessing the adequacy of data interconnections across different domains in meeting practical needs.

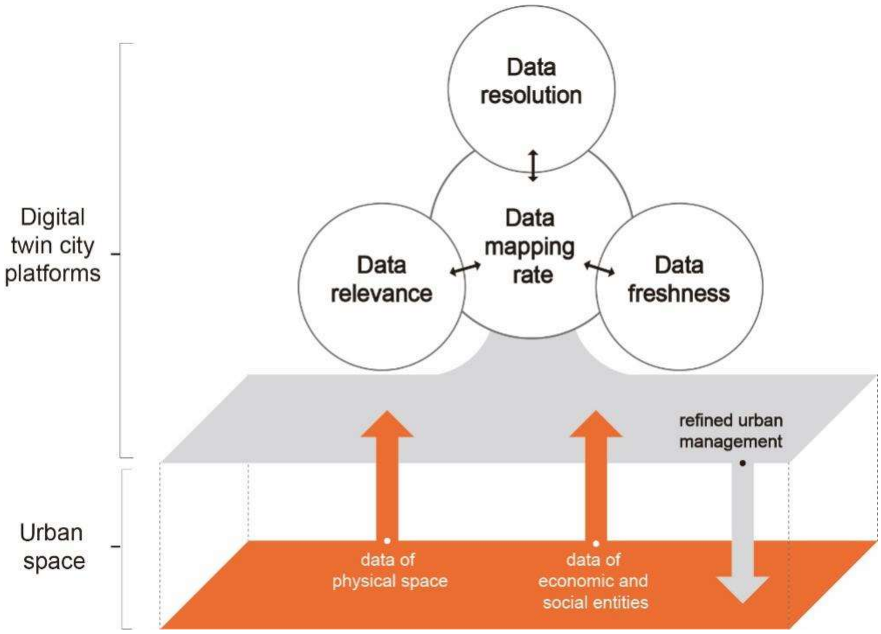

Fig. 1 Data mapping rate towards digital twin city platform

For a quantitative assessment, the mapping rate evaluation model should be predicated on uniform and systematic urban data. Thus, investigating the data dimension structure of digital twin city platforms is essential. This study examines extensive urban data, categorized into four types. On one hand, it emphasizes data on the foundational physical elements of cities, namely "Spatial Carriers" and "Urban Infrastructure." On the other hand, in contrast to traditional projects, there is an increased emphasis on systematically collecting and integrating Socioeconomic Data, including data from "Social Entities" (e.g., residents, economic entities, governments, social organizations) and "Urban Datastreams" (e.g., human traffic, transportation flow, capital movement). These types are hierarchically categorized into

"Major Categories," "Subcategories," and "Minor Categories," to capture the intricate array of data resources imperative for multi-entity and multi-structured urban management. This data system is widely adaptable for assessments across diverse digital twin cities, offering a standard benchmark to compute data mapping rates efficiently.

3 Methodology

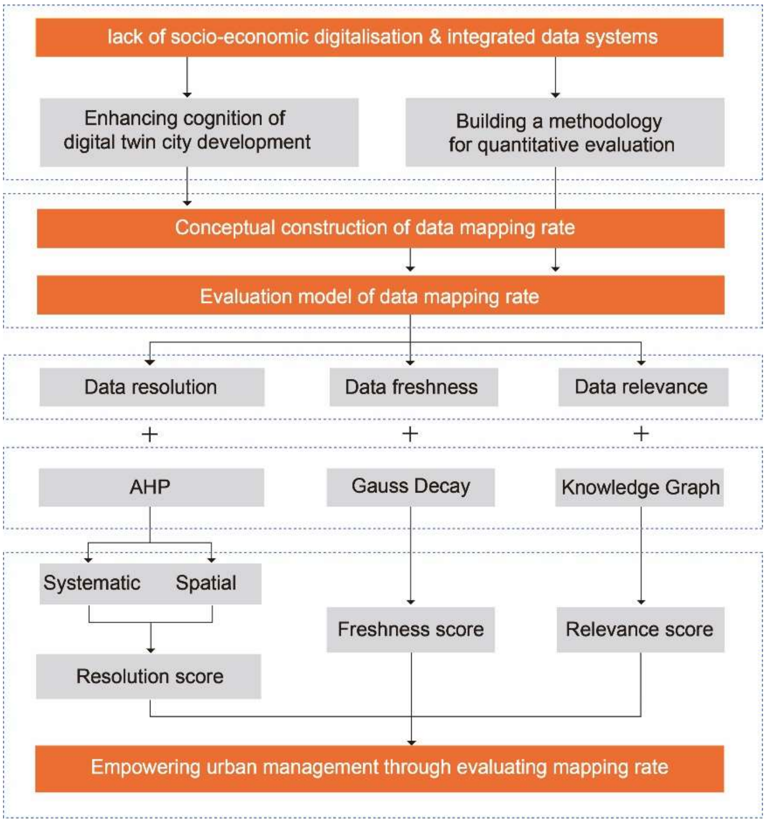

Fig. 2 Research framework

To effectively measure data mapping rates, this study integrates methods such as the Analytic Hierarchy Process (AHP), Gaussian functions, and knowledge graphs to establish an evaluation model centered on data mapping rates. Using the Huamu digital twin platform as a case study, we conducted a detailed measurement of the previously unmeasurable data construction status for fine-grained evaluation. The socioeconomic data related to privacy were collected by the Huamu government and subsequently underwent appropriate security measures before analysis. This method will guide the sustainable development of urban digital twin platforms and precisely empower refined urban management (see Figure 2).

3.1 Measurement of Data Resolution

Data resolution comprises systematic resolution and spatial resolution. System resolution is a measure of the completeness of data collection. We convert the coverage of data dimensions within the digital twin platform into the total area (S) of resolution units, as shown in the following formula:

$$\mathrm{S} = \sum_i L_i \times W_i \times k_i \tag{1}$$

Here, W_i represents the type weight of the different dimensions, characterizing the differences in significance of different data types. L_i represents the depth weight of the data level, differentiating between different levels of granularity. k_i represents the data integrity in each dimension, indicating the comprehensiveness of the collection of fields in a specific dimension.

For W_i, we utilize the AHP, a quantitative method for dealing with complex decisions (16). Weights were automatically calculated based on the significance of various data types in the analysis based on the judgments of three domain experts. For L_i, we assign exponentially increasing weights of 1, 2, and 4 to "major categories," "subcategories," and "minor categories," respectively, reflecting the rising costs of detailed data collection. For k_i, we differentiate between "mandatory fields" (M items, constituting 60%) and "bonus fields" (N items, making up 40%) across each dimension. The field collection rate is defined as the proportion of actually collected fields to the total required fields, represented by ∂^k.

$$k_i = 0.6\left(\frac{\sum_M^k \partial^k}{M}\right) + 0.4\left(\frac{\sum_N^k \partial^k}{N}\right) \tag{2}$$

Based on this framework, attaining a k_i of 1 across all data dimensions indicates the collection of comprehensive data, equivalent to 100 points. The scoring for systematic resolution is determined by comparing this ideal total area with the actual total area of resolution units for the specific data platform.

The methodology for spatial resolution measurement is similar to systematic resolution but targets to modeling three-dimensional virtual cities. It is based on the "Technical Guidelines for the Urban Information Model (CIM) Basic Platform", ensuring its applicability to real projects. The calculation of W_i and L_i for spatial resolution is identical to systematical resolution, with the difference being the binary evaluation (0 or 1) of k_i.

3.2 Measurement of Data Freshness

Data freshness, within the framework of data maintenance, refers to how frequently data is refreshed. Recognizing the diverse requirements for update frequencies among various data sets in urban management, it is crucial to determine the suitable update intervals for specific datasets. This study has developed a framework that balances the need and cost of updates, resulting in the establishment of five update frequency categories: "Daily," "Next-Day," "Monthly," "Semiannual," and "Annual," which signify the advised refresh cycles. The decay curves of data freshness were modelled using Gaussian functions. This assumes that data peaks in freshness upon update and decays to half at the suggested update cycle, thus determining the function forms for varying data. This method provided a formula for the decay of data freshness and its associated coefficients α_n.

$$f_n(x) = e^{\alpha_n x^2} \tag{3}$$

We performed an extensive examination of the foundational fields within the data system, leading to the development of a "Data-Field-Update Cycle" mapping table. Owing to the hierarchical tree structure of the data system, it is possible to aggregate the freshness scores from each level to determine the freshness of the immediate higher level. Consequently, the overall data freshness of the platform can be quantified using the subsequent formula:

$$X_{\text{n-1}} = \sum_i W_i \times X_n \tag{4}$$

X_n signifies the composite freshness of data across multiple dimensions at the nth level; when referring to the base-level data, X_n corresponds to the freshness of fields. Furthermore, considering the differential significance of updating diverse data types within urban management, this study introduces $W_i$as weighted values for each dimension using the AHP.

3.3 Measurement of Data Relevance

The evaluation of data relevance aims to ascertain the extent to which data relevant to urban management scenarios are interconnected. To measure these connections and the structure of the data network, we utilize Gephi, a software for network analysis, to apply the methods of knowledge graphs (4). This analysis examines the correlations among data and the structural characteristics of the network, facilitating network visualization. Furthermore, a fuzzy matching algorithm is employed to compare fields and estimate the probability of their association with the same entity.

This approach enables a comprehensive assessment across various service scenarios and mitigates interference caused by discrepancies in field naming.

Specifically, the construction of the data network and the calculation of data relevance are performed in Gephi. Within a complex network, different types of urban data are represented as nodes, and the field co-occurrence—determined through fuzzy matching—acts as the edge-weight connecting these nodes. The score of data relevance is quantified using the average degree (8).

$$C(i) = \sum_{i\neq j, i\neq q, j\neq q} \frac{\delta_{jq}(i)}{\delta_{jq}} \tag{5}$$

In this formula, $\delta_{jq}(i)$ represents the edge weight, which corresponds to the frequency of field co-occurrence between tables, while δ_{jq} denotes the number of nodes in the network.

4 Results

The Huamu sub-district, located in the Pudong New Area of Shanghai, is characterized by its high population density and complex management tasks, making top-down management challenging. In the context of Shanghai government's digital transformation, Huamu initiated the construction of a digital twin city platform starting from 2020. Currently, the Huamu Street sub-district has established a visualization platform along with a corresponding data collection and updating system, which is gradually being applied in grassroots management.

This study evaluated the data mapping rate of the Huamu digital twin city platform, as shown in Figure 3. The platform achieved an overall data resolution score of 39.7. The systematic resolution score was 41.5, indicating a favorable performance. However, data collection for "Urban Datastreams" exhibited significant deficiencies. The spatial resolution score, documented at 37.1, pointed to a mediocre performance in modeling completeness and detail. The data freshness score was 34.1, which was considered moderate. Despite the relatively high freshness of "Spatial Carriers," both "Urban Datastreams" and "Urban Infrastructure" were hampered by inadequate data updating mechanisms. The data relevance score was a disappointing 13.9, with "Social Entities" and "Spatial Carriers" showing strong associations, in contrast to "Urban Infrastructure," which demonstrated the weakest connectivity with other dimensions.

For data resolution, efforts to refine the Huamu digital twin platform must prioritize the enrichment of data pertaining to "Urban Infrastructure" and "Social Entities," and pivot towards establishing effective channels for collecting "Urban Datastreams" data in the next phase. Modeling initiatives should also incorporate elements such as "Underground Spaces" and "Municipal Pipelines". For data freshness, a concentrated effort is needed to improve the recency of "Urban Datastreams" data and to implement a robust updating process for "Urban Infrastructure," with a particular focus on synchronizing data with frequent update requirements. For data relevance, it is imperative to enhance the database's completeness, encompassing more pivotal data nodes and fields. Moreover, in the process of data standardization, establishing uniform naming conventions is critical to facilitate more efficient data matching processes.

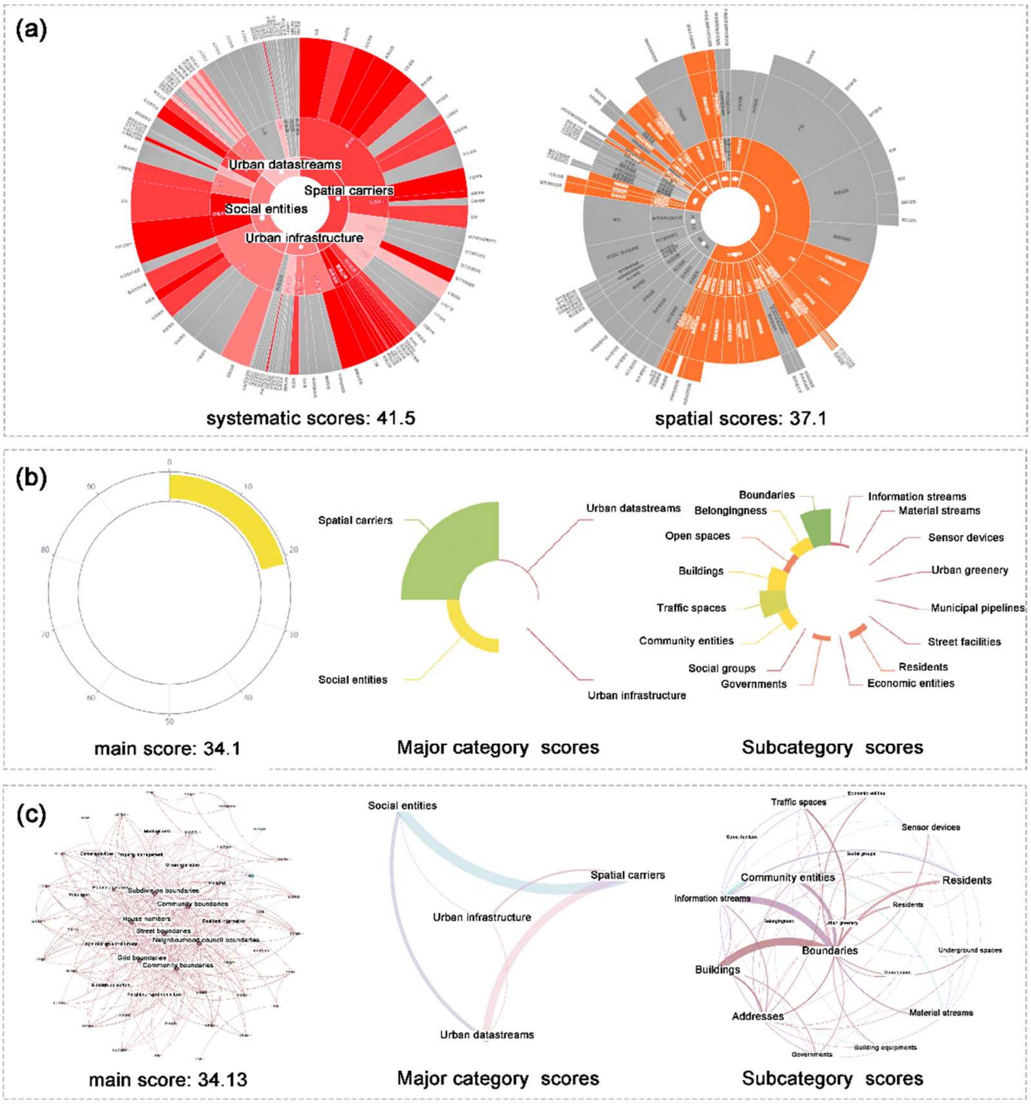

Fig. 3 Data mapping rate scores of Huamu digital twin city platform

5 Discussions and Conclusions

This study introduces a research framework aimed at evaluating the mapping rate of digital twin city platforms, fulfilling the need for refined urban management solutions. It addresses specific challenges encountered in the development of digital twin platforms by emphasizing the importance of data collection completeness, update timeliness, and network interconnectedness to define the concept of data mapping rate. Moreover, we have developed a data evaluation model focused on the data mapping rate, facilitating the previously unquantifiable assessment of digital twin platform progress. This framework, through the construction of a hierarchical data system and a combination of a series of usable and efficient methods, is widely applicable to various digital platforms. By emphasizing data freshness and relevance, it encourages relevant practices to emphasize the dynamic interrelations and effectiveness of data. Consequently, this enhances the responsiveness and adaptability of urban management, effectively addressing rapidly changing urban environments.

Furthermore, this research outlines an effective strategy for the systematization, diversification, and standardization of urban digital management practices. Unlike traditional top-down management models (12), the digital twin city management approach, predicated on the mapping rate, heralds a bottom-up system construction. Starting from the community level as the fundamental unit, it is possible to establish a higher-level digital management network, paving the way for the development of an advanced and comprehensive smart system. This evolution promises to elevate urban management and service efficiency.

However, the study acknowledges certain limitations. First, the application of the AHP introduces subjectivity in the determination of periods and weights for various indicators, although it still provides a relatively reliable benchmark for the calculation of data mapping rates. Future research may explore the use of stated preference methods for optimization. Second, the objective of measuring the mapping rate is not to indiscriminately strive for higher values. Instead, it seeks to identify an optimal range that balances management improvement and input costs, tailored to the project's specific needs. Nonetheless, given the current underdeveloped data infrastructure of many platforms, any advancement in data resolution is deemed beneficial.

6 References

1. Al-Sehrawy R, Kumar B, Watson R. A digital twin uses classification system for urban planning & city infrastructure management. Journal of Information Technology in Construction. 2021;26:832-362.
2. Allam Z, Dhunny ZA. On big data, artificial intelligence and smart cities. Cities. 2019;89:80-91.

3. Chen M, Liu W, Lu D, Chen H, Ye C. Progress of China's new-type urbanization construction since 2014: A preliminary assessment. Cities. 2018;78:180-93.
4. Chen X, Jia S, Xiang Y. A review: Knowledge reasoning over knowledge graph. Expert Systems with Applications. 2020;141:112948.
5. Dani AAH, Supangkat SH, Lubis FF, Nugraha IGBB, Kinanda R, Rizkia I. Development of a Smart City Platform Based on Digital Twin Technology for Monitoring and Supporting Decision-Making. Sustainability. 2023;15(18):14002.
6. Engin Z, van Dijk J, Lan T, Longley PA, Treleaven P, Batty M, et al. Data-driven urban management: Mapping the landscape. Journal of Urban Management. 2020;9(2):140-50.
7. Hashem IAT, Chang V, Anuar NB, Adewole K, Yaqoob I, Gani A, et al. The role of big data in smart city. International Journal of information management. 2016;36(5):748-58.
8. Hussain S, Muhammad L, Yakubu A. Mining social media and DBpedia data using Gephi and R. Journal of Applied Computer Science & Mathematics. 2018;12(1):14-20.
9. Ivanov S, Nikolskaya K, Radchenko G, Sokolinsky L, Zymbler M, editors. Digital twin of city: Concept overview. 2020 Global Smart Industry Conference (GloSIC); 2020: IEEE.
10. Klauser FR, Albrechtslund A. From self-tracking to smart urban infrastructures: Towards an interdisciplinary research agenda on Big Data. Surveillance & Society. 2014;12(2):273-86.
11. Lu Q, Parlikad AK, Woodall P, Don Ranasinghe G, Xie X, Liang Z, et al. Developing a digital twin at building and city levels: Case study of West Cambridge campus. Journal of Management in Engineering. 2020;36(3):05020004.
12. Pissourios I, policy. Top-down and bottom-up urban and regional planning: Towards a framework for the use of planning standards. European spatial research. 2014;21(1):83-99.
13. Rasheed A, San O, Kvamsdal T. Digital twin: Values, challenges and enablers from a modeling perspective. Ieee Access. 2020;8:21980-2012.
14. Schrotter G, Hürzeler C. The digital twin of the city of Zurich for urban planning. PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science. 2020;88(1):99-112.
15. Shahat E, Hyun CT, Yeom C. City digital twin potentials: A review and research agenda. Sustainability. 2021;13(6):3386.
16. Vaidya OS, Kumar S. Analytic hierarchy process: An overview of applications. European Journal of operational research. 2006;169(1):1-29.
17. Weil C, Bibri SE, Longchamp R, Golay F, Alahi A. A Systemic Review of Urban Digital Twin Challenges, and Perspectives for Sustainable Smart Cities. Sustainable Cities and Society. 2023:104862.
18. White G, Zink A, Codecá L, Clarke S. A digital twin smart city for citizen feedback. Cities. 2021;110:103064.
19. Yossef Ravid B, Aharon-Gutman M. The social digital twin: the social turn in the field of smart cities. Environment and Planning B-Urban Analytics and City Science. 2023;50(6):1455-70.