Ubiquitous geographic information for building digital twins of geographic environments

Guonian Lü,[1,2,3,*] Songshan Yue,[1,2,3,*] Zhaoyuan Yu,[1,2,3] Min Chen,[1,2,3] and Liangcheng Zhou[1,2,3]

[1]Key Laboratory of Virtual Geographic Environment (Ministry of Education of PRC), Nanjing Normal University, Nanjing 210023, China
[2]State Key Laboratory Cultivation Base of Geographical Environment Evolution (Jiangsu Province), Nanjing 210023, China
[3]Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China
*Correspondence: gnlu@njnu.edu.cn (G. L.); yuesongshan@njnu.edu.cn (S. Y.)

Received: June 12, 2023; Accepted: September 4, 2023; Published Online: September 4, 2023; https://doi.org/10.59717/j.xinn-geo.2023.100023

Citation: Lü G., Yue S., Yu Z., et al., (2023). Ubiquitous geographic information for building digital twins of geographic environments. The Innovation Geoscience 1(2), 100023.

With the rapid development of information and communications technology, we are now living in a triadic world where physical, social and information spaces are closely interconnected via matter transfer, energy conversion, and information dissemination[1]. Physical space provides the essential elements of air, water, land, and life for the human beings. Moreover, within the realm of social space, we reform the physical space through diverse economic, political, and cultural activities. Information space has the capability to mirror the elements of the other two spaces and influence the natural processes and human activities that occurring within them. Many global and local challenges we encounter, such as climate change, food crises, water security and unsustainable development, are related to the interactions among these three spaces. To solve these problems, the triadic world needs to be modeled effectively. Building digital twins of geographic environments is considered as a prominent idea[2], but most existing digital twins only reflect the physical space and focus on technical factors. The social and information spaces are important for human adaptation and world change , yet they are overlooked to a large extent. To overcome this limitation, it is necessary to improve the framework for digital twins and include ubiquitous geographic information from all three spaces.

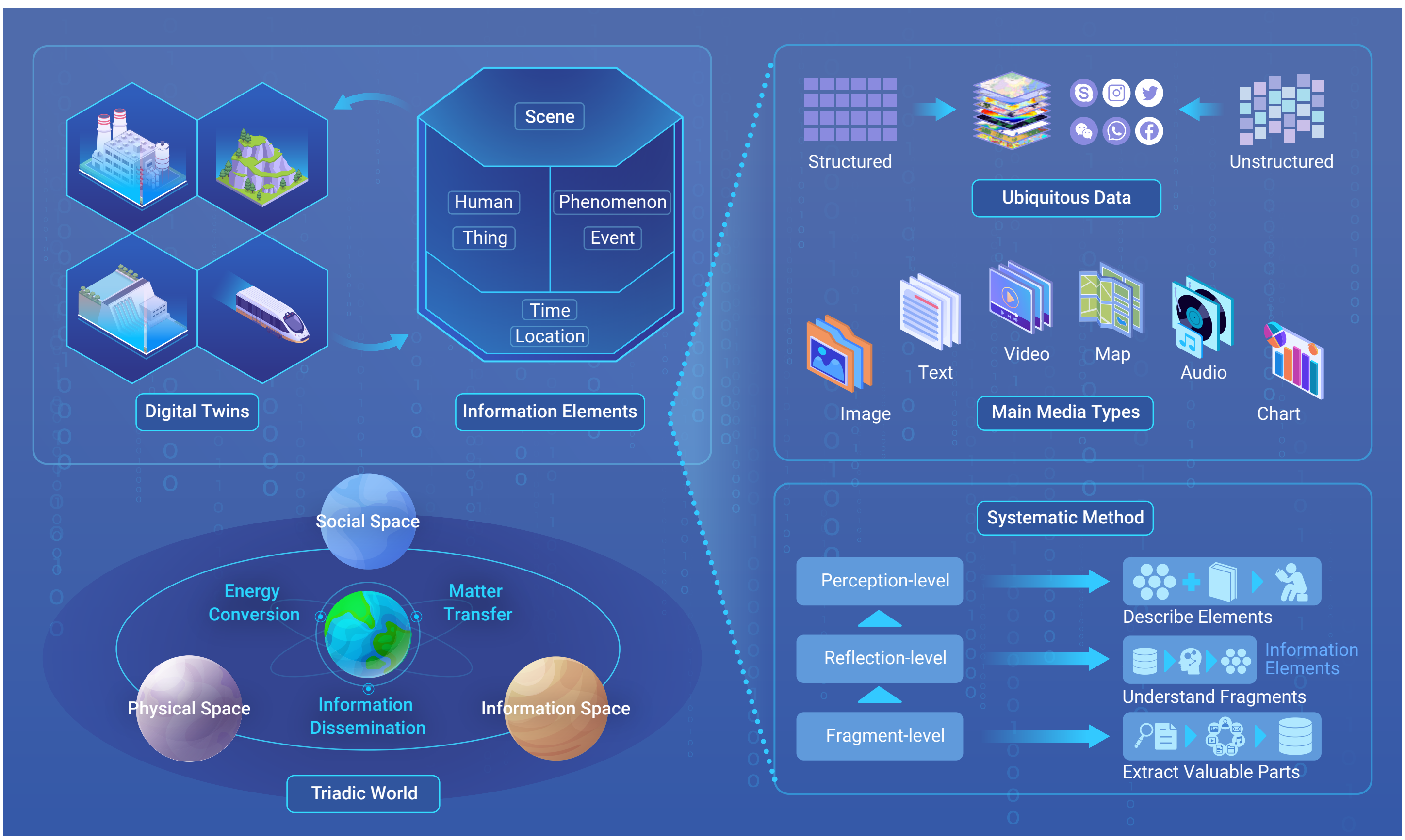

Figure 1. Interpreting information elements from ubiquitous data for digital twins

ELEMENTS OF A DIGITAL TWIN OF THE GEOGRAPHIC ENVIRONMENT

Digital twins originated from the industrial sector, where they were used to manage and maintain equipment in factories and workshops. In the geographic information community, digital twins are mainly technical tools used to create digital representations of the physical world, such as digital twin cities, dams and watersheds[3]. The term "twins" implies that computer models can "mirror" physical geo-objects and receive dynamic observations from Internet of Things (IoT) devices. However, critics have raised the following question: Since the idea of building digital models of the physical world is not new (e.g., digital Earth and digital cities), what are the novel features of digital twins?

We believe that the key difference is in the framework of representing the geographic environment. Traditional digital models and existing digital twins tend to use a geometric framework to organize information. However, the geographic environment is a huge and complex system that involves many elements and nonlinear interactions among these elements. A geometric framework, which relies on coordinates with a specific resolution and reference, restricts the perspective from which the geographical environment can be modeled. Although it is easy to visualize and store related information in databases, the geometric framework is unsuitable for understanding the real world in many other ways, such as in the context of a person's life or a rainstorm process. A digital twin should support different perspectives, and the corresponding framework should be customized by combining information elements.

We identified seven types of information elements to describe the real world: time, location, human, thing, event, phenomenon and scene. There might be other terms that can describe each of the seven element types, and their core connotations must remain consistent. Time elements indicate when, location elements show where, human elements define who, thing elements show what, event elements represent occurrence, phenomenon elements stand for recurrence, and scene elements provide context. Time and location elements are essential for supporting the spatiotemporal background of a digital twin. Human and thing elements are referring to geo-objects that exist in the real world, and distinguished based on whether the object has human characteristics or not. Event and phenomenon elements are related to dynamic processes that occur in the real world. They are distinguished based on whether the process has specific time and location boundaries. Scene elements comprise a group of aforementioned elements with contextual relationships that form our understanding of the real world. A scene element can contain multiple subscenes. The combination of the seven types of elements forms the new framework of the digital twins of the geographic environment. In addition, in the location-based framework (a geometric framework), each type of element and combinations of elements can be customized to build a digital twin that comprehensively depict the real world (Figure 1).

INTERPRETING INFORMATION ELEMENTS FROM UBIQUITOUS DATA

Information elements associated with the seven types of features can be obtained from different data sources in the real world. The process of obtaining information elements from data involves different research processes, such as environmental sensing, data mining, and knowledge discovery. Many efforts have been made to extract useful information from data, and many advances have been made, especially in the field of big data mining[4]. However, there are still two questions to be answered: (1) How many data sources exist in the world? and (2) How many methods can be used to obtain information elements from different data sources? These questions are important because building a complete digital twin of the geographic environment requires a "big vision" to combine information from "big data"[5]. It is too difficult for one discipline to provide such a big vision. The integration of multidisciplinary knowledge and heterogeneous data sources is needed to build digital twins. Even a digital twin scaled to a small area require substantial amount of data. More importantly, a digital twin can be improved by continually adding information elements. Knowing the various types of available data sources and methods could help users build and improve digital twins in a systematic way.

All data, no matter how it is collected or what information it encompasses, can be organized into these six media types: text data in natural languages, image data in pixels, audio data in signals, video data in frames, map data in symbols, and chart data in shapes. For example, scientific data organized as images (e.g., remote sensing images, grid-based elevations and land uses) are part of the image data group. The corresponding structure-matching rule applies to other media types. Most scientific data with specific formats are part of the text data group, as a data format can also function as a kind of language that describes the associated content. Unlike scientific datasets with clear structures, data from the general public are more flexible and offer valuable content to understand the real world. Twitter, Weibo, Instagram, TikTok, We-media, and other software packages and apps serve as important data sources. Data from both academia and the public can be divided into six ubiquitous data types. Each of the six data types has subtypes based on different criteria, such as domain, language, color, resolution, and device. A digital twin can be built in a systematic way by selecting data sources from the hierarchical classification tree of ubiquitous data.

In terms of the methods used to obtain information elements from ubiquitous data, rule-based, model-based, machine learning, and deep learning methods have been widely explored and developed. However, most of these methods are self-evolved and focus on interpreting specific features of certain data types. A system that links individual methods with the hierarchical classification tree of ubiquitous data can assist users in discovering and utilizing suitable methods more conveniently. We propose a three-level system for extracting information elements from ubiquitous data. In the first level, fragmented materials are identified. In the second level, the fragmented materials are classified into orderly information elements, and in the third level, a bridge between information elements and holistic perception is constructed (Figure 1).

Fragment-level methods are used to extract valuable parts of original datasets. The results are normally presented as they appear in their original state (e.g., some words in a sentence, an object's boundary in an image, and certain frames in a video). Reflection-level methods are used to understand what the fragments are and convert them into the seven types of information elements. The information elements for a formal structure are generated, and information types are assigned to the original fragments. Perception-level methods are used to describe in detail the content of information elements based on contextual understandings. External knowledge might be incorporated to help enrich information elements from different perspectives. Taking a place name in text data as an example, it can be extracted as certain words (fragment-level methods), parsed as location elements (reflection-level methods), and then interpreted based on its coordinates, historical evolution, administrative structure, and other related content (perception-level methods).

SUMMARY

Digital twins of geographic environments can help us model the real world and make it better. However, the existing digital twins focus on the physical space rather than physical-social-information triadic space. This limits their ability to depict the geographic environment from diverse and holistic perspectives. Through the construction of a framework with basic types of information elements (i.e. time, location, human, thing, event, phenomenon and scene) and the classification of ubiquitous data (i.e. text, image, audio, video, map and chart), a system can be formed to help interpret ubiquitous data and provide various information elements for digital twins. These seven types of information elements can be combined to form various digital twins, and the theme of a digital twin determines which elements should be included. Digital twins within the framework of ubiquitous geographic information can be comprehensive and offer different views of the real world.

REFERENCES

1. Lü, G., Yuan, L., and Yu, Z. (2022). Information Geography: a new fulcrum of geographic ternary world. Sci. China Earth Sci. 65, 383-386. DOI: 10.1007/s11430-021-9859-9.
2. Caldarelli, G., Arcaute, E., Barthelemy, M., et al. (2023). The role of complexity for digital twins of cities. Nat Comput Sci 3, 374-381. DOI: 10.1038/s43588-023-00431-4.
3. Bauer, P., Stevens, B. and Hazeleger, W. (2021). A digital twin of Earth for the green transition. Nat. Clim. Chang. 11, 80-83. DOI: 10.1038/s41558-021-00986-y.
4. Li, X., Feng, M., Ran, Y., et al. (2023). Big Data in Earth system science and progress towards a digital twin. Nat Rev Earth Environ 4, 319-332. DOI: 10.1038/s43017-023-00409-w.
5. Tzachor, A., Sabri, S., Richards, C. E., et al. (2022). Potential and limitations of digital twins to achieve the Sustainable Development Goals. Nat Sustain 5, 822-829. DOI: 10.1038/s41893-022-00923-7.

FUNDING AND ACKNOWLEDGMENTS

The work is supported by the National Natural Science Foundation of China (Nos. 41930648 and 42171406). The founders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

DECLARATION OF INTERESTS

Min Chen is an Editorial Board member of The Innovation Geoscience and was blinded from reviewing or making final decisions on the manuscript. Peer review was handled independently of this member and their research group. The other authors declare no conflicts of interest.