Journal of Earth Science, Vol. 34, No. 5, p. 1323−1327, October 2023 ISSN 1674-487X
Printed in China
https://doi.org/10.1007/s12583-023-1930-1

Perspective

Knowledge System, Ontology, and Knowledge Graph of the Deep-Time Digital Earth (DDE): Progress and Perspective

Xiumian Hu[*1], **Yiwei Xu**[2], **Xiaogang Ma**[3], **Yunqiang Zhu**[4], **Chao Ma**[5], **Chao Li**[6], **Hairong Lü**[7], **Xinbing Wang**[8], **Chenghu Zhou**[4], **Chengshan Wang**[9]

1. *School of Earth Sciences and Engineering, Nanjing University, Nanjing* 210023, *China*
2. *State Key Laboratory of Palaeobiology and Stratigraphy, Nanjing Institute of Geology and Palaeontology and Center for Excellence in Life and Paleoenvironment, Chinese Academy of Sciences, Nanjing* 210008, *China*
3. *Department of Computer Science, University of Idaho, Moscow ID* 83844, *USA*
4. *Institute of Geographical Science and Natural Resources, Chinese Academy of Sciences, Beijing* 100101, *China*
5. *State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation & Institute of Sedimentary Geology & College of Computer Science and Cyber Security, Chengdu University of Technology, Chengdu* 610059, *China*
6. *IUGS Deep-Time Digital Earth Research Center of Excellence, Suzhou* 215101, *China*
7. *Department of Automation, Tsinghua University, Beijing* 100084, *China*
8. *School of Electronic, Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai* 200240, *China*
9. *State Key Laboratory of Biogeology and Environmental Geology, China University of Geosciences, Beijing* 100083, *China*

THE USE OF KNOWLEDGE GRAPH IN NATURAL SCIENCE

Knowledge graph is a field of Artificial Intelligence (AI) that aims to represent knowledge in the form of graphs, consisting of nodes and edges which represent entities and relationships between nodes respectively (Aidan et al., 2022). Although the knowledge graph was popularized recently due to use of this idea in Google's search engine in 2012 (Amit, 2012), its root can be traced back to the emergence of the Semantic Web as well as earlier works in ontology (Aggarwal, 2021). These semantic networks are graphical representations of relationships among entities, which can be used to support propositional or first-order logic in traditional knowledge bases and natural language processing applications (Aggarwal, 2021).

Knowledge graphs consist of two main components: ontology and facts. The ontology comprises concepts and their relationships, while facts consist of entities and their relationships (Fig. 1). Knowledge graphs have been utilized in scientific research across two primary areas in the past decades. Firstly, knowledge graph was considered as a type of knowledge base that can effectively function as knowledge service, including information retrieval, knowledge quiz, knowledge reasoning, and visualization display. A notable example demonstrating this application is the AceKG system developed by Wang et al. (2018). The second aspect pertains to the ontological level, where domain-specific terminologies are transformed into the standardized vocabularies through structural representation. This fosters the development of machine-readable knowledge systems, integrating diverse conceptual information from various datum systems. Consequently, disparate datum sources can be harmonized and consolidated through the utilization of ontologies, ensuring interoperability and facilitating knowledge integration. The Gene ontology and other biomedical domain ontologies (e.g., ULMS) are used in this way, and have significantly enhanced integration of diverse types of structured data and enabling the development of tools for mining unstructured text data (Gene Ontology Consortium, 2019; Lexical Systems Group, 2018).

The 21st century has witnessed an exponential growth in scientific data, driven by advancements in technology. The field of geosciences, in particular, has amassed a substantial volume of data, propelling geoscientists into the era of big data (Ma X et al., 2023; Wang et al., 2021). However, data in the field of geoscience are stored in databases with diverse datum standards or unstructured literature texts. Challenges arise due to inconsistent terminologies, unclear sharing mechanisms, and semantic heterogeneity, which impede the effective utilization of data. As a result, the advantages of big data are less frequently applied in the geoscience (Zhu et al., 2023b). The integration, mining, and sharing of both structured and unstructured data inevitably require the support from knowledge of the geoscience, therefore the construction of knowledge graph serves as the foundation for big data-driven research in geosciences. The geoscience research has been entering a significant transitional period with the establishment of a new knowledge system as the core and with the drive of big data as the means (Zhu et al., 2023c; Zhou et al., 2021).

There has been notable progress in the development of technologies for constructing generalized knowledge graphs in the past years, several generalized knowledge graphs have been successfully created (e.g., Wikidata, YAGO etc.) and applied within the field of AI. The construction methods of knowledge graphs of geoscience are still poorly explored, with

*Corresponding author: huxm@nju.edu.cn
© China University of Geosciences (Wuhan) and Springer-Verlag GmbH Germany, Part of Springer Nature 2023

Manuscript received August 29, 2023.
Manuscript accepted September 15, 2023.

Hu Xiumian, Xu Yiwei, Ma Xiaogang, Zhu Yunqiang, Ma Chao, Li Chao, Lü Hairong, Wang Xinbing, Zhou Chenghu, Wang Chengshan , 2023. Knowledge System, Ontology, and Knowledge Graph of the Deep-Time Digital Earth (DDE): Progress and Perspective. *Journal of Earth Science*, 34(5): 1323–1327. https://doi.org/10.1007/s12583-023-1930-1. http://en.earth-science.net

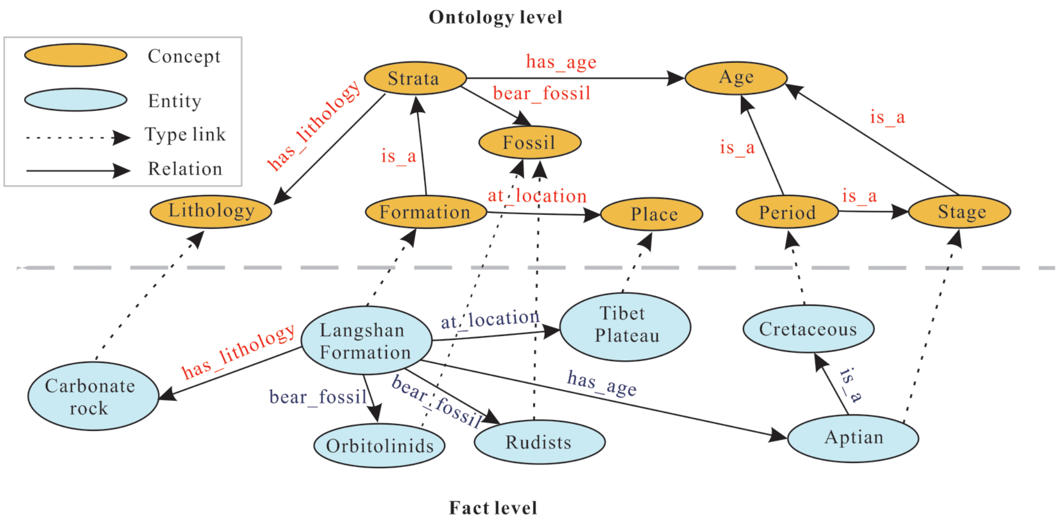

Figure 1. An example of two-level of knowledge graph.

limited studies specifically targeting this domain. Existing knowledge graphs in geoscience can be grouped into two categories. Group 1 comprises generalized knowledge graphs of earth sciences, such as GeoSciML (Simons et al., 2006) and SWEET ontology (Buttigieg et al., 2018) that have been specifically constructed to facilitate knowledge sharing within the geoscience domain. These graphs are somewhat similar to the top ontologies designed by domain experts. As a consequence, they contain a limited number of concepts. For instance, the GeoSciML and SWEET ontologies consist of a total of 1 772 and 4 533 concepts, respectively (Zhu et al., 2023a). Therefore, they cannot meet the requirements for handling the extensive knowledge within the entire geoscience domain. Group 2 represents knowledge graphs for specific fields within geoscience. Some of these knowledge graphs are designed to integrate disparate databases, such as Linked Earth, which focuses on paleoclimatic data integration (Khider et al., 2019). Others serve as standardized terminologies for specific domains, including geological time scale (Ma and Fox, 2013), sedimentary rock (Abel et al., 2012) and other fields. However, these knowledge graphs are constructed using various languages (e. g., OWL, RDF, Turtle, etc.) and tools (Protégé, Neo4j, etc.). Additionally, the relationships between different graphs are often unclear, posing challenges for the effective integration and reuse of them.

PRESENT WORK

Considering the current state of knowledge graph construction in the field of geoscience, the deep-time digital Earth program (DDE) plans to construct a comprehensive knowledge graph of the geoscience by establishing a unified representation model. This endeavor aims to facilitate the establishment of standardized terminology and the integration of knowledge within the field of geoscience (Wang et al., 2021; Zhou et al., 2021).

To achieve this objective, the DDE knowledge graph follows a three-stage construction plan (Fig. 2). The initial stage focuses on the development of the knowledge system, commonly known as the specialist dictionary (Fig. 3a). In this stage, domain experts in geoscience enumerate key terms within the disciplinary domain, providing precise definitions and referencing relevant literature for each term, thereby establishing a specialist dictionary. Ensuring accurate and standardized definitions of terminology is vitally import for the DDE, as it facilitates the harmonization of deep-time earth data and the creation of a knowledge engine that supports abductive exploration of Earth's evolution.

The terms within the specialist dictionary are initially independent from one another or limited hierarchical relations. Consequently, the second stage of the knowledge graph construction involves establishing relationships and properties between among terms, resulting in the formation of ontologies (Fig. 3b). The relationships within the ontology enable to support automated reasoning and to facilitate the integration of different concepts within a unified framework. The ontologies of DDE are classified into two types: foundational ontology, encompassing commonly used ontologies across geoscience disciplines like spatial ontology (Wang S et al., 2023) and geological time ontology (Ma C et al., 2023; Ma and Fox, 2013), and domain-specific ontology, consisting of specialized content from various sub-disciplines in geosciences. Currently, 20 domain-specific ontologies have been preliminarily developed within the framework of the DDE (Table 1). These ontologies will be expanded from literature corpora by using techniques such as natural language processing and text analytics. While the adoption of community-level foundational ontologies reflects the top-down approach, the completion of the DDE knowledge graph will be achieved by augmenting facts (entities and relationships) through bottom-up methods that leverage the literature data. The combination facts and ontology of geoscience facilitates the reusability of data, the discovery of relevant information, and the generation of new novel insights through logical reasoning.

To enable an open access to the expert-built ontologies, the DDE has developed the Geoscience Knowledge Graph Collaborative Editor (Fig. 4; Zhu et al., 2023b; Shi et al., 2020). Over the four-year period (2019–2023), the DDE knowledge graph has achieved its first-stage construction goals on this

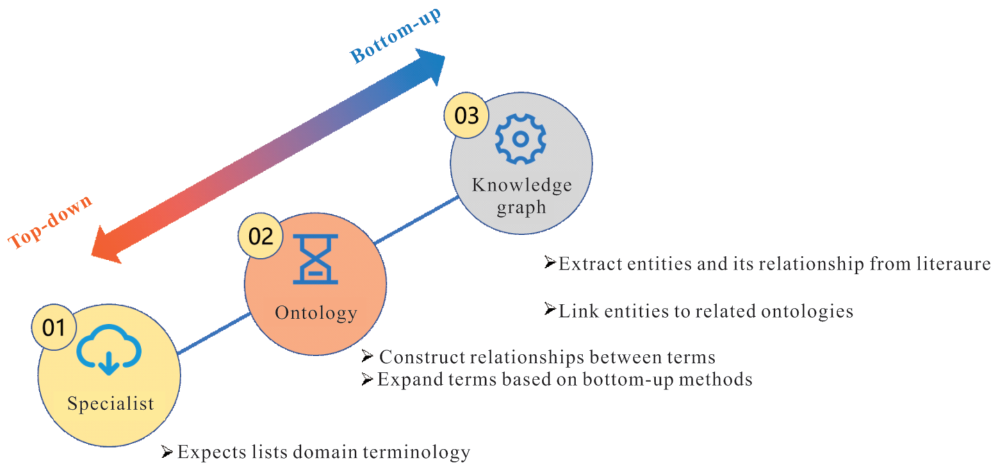

Figure 2. Construction workflow for DDE knowledge graph.

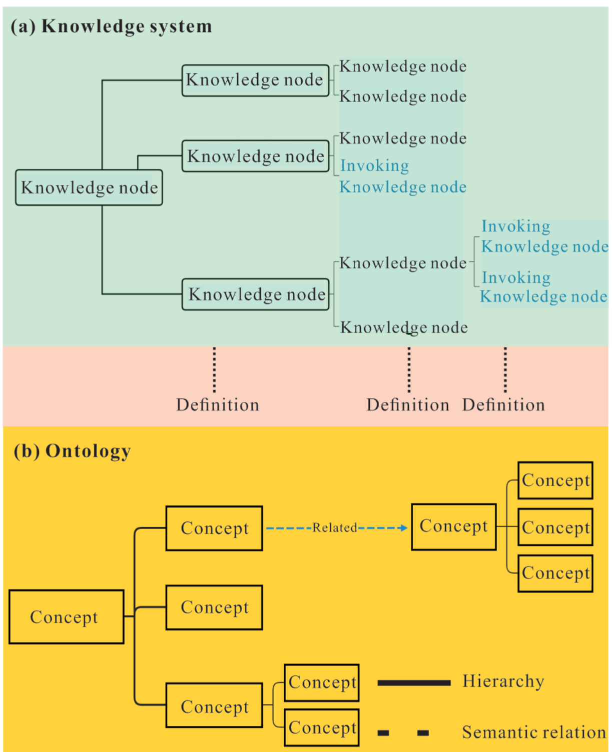

Figure 3. Frameworks of (a) knowledge system and (b) ontology of DDE.

platform and has now entered the second-stage ontology construction phase. Currently, the DDE ontology comprises 61 937 nodes and 62 610 relationships (Table 1). In terms of both scale and content, the DDE knowledge graph represents the most comprehensive and versatile knowledge graph in geoscience.

Although the work on the knowledge graph is ongoing, the DDE ontologies have already exhibited practical value in scientific research. The ontology's reasoning capabilities have been employed to automatically identify sedimentary environments, specifically river facies and carbonate platforms (Wang H et al., 2023; Zhang L et al., 2023). The integrated terminology function has facilitated to mine the global data of stromatolites (Zhang X B et al., 2023), while the retrieval function has been utilized for managing stratigraphic database (Xu J L et al., 2023). In addition to the current 20 domain ontologies within the DDE, several knowledge graphs have been developed to meet specific scientific needs. These include the carbonate rocks (Xu et al., 2023b), paleobiogeography and climate paleogeography (Yu et al., 2023; Zhang L N et al., 2023), petroleum exploration (Tang et al., 2023), tectonic geomorphology (Xi et al., 2023), geothermal (Chen et al., 2023) and others. Multiple studies (Ma K et al., 2023; Parson et al., 2023; Qiu et al., 2023a, b) have proposed state-of-the-art techniques and standards to form the foundation for the DDE knowledge graph.

Table 1 DDE domain ontology

Domain	Nodes	Domain	Nodes
Paleontology	24 936	Mathematical geoscience	765
Stratigraphy	1 268	Geomagnetism paleomagnetism	2 434
Sedimentology	2 675	Engineering geology	3 446
Paleogeography	2 831	Petroleum geology	2 390
Mineralogy	5 665	Geothermic	907
Igneous petrology	1 670	Hydrogeology	806
Metamorphic petrology	1 028	Geomorphology	1 958
Mineral deposit	605	Surficial geochemistry	3 690
Structure geology and tectonics	1 291	Geological mapping	1 990
Geochronology	424		

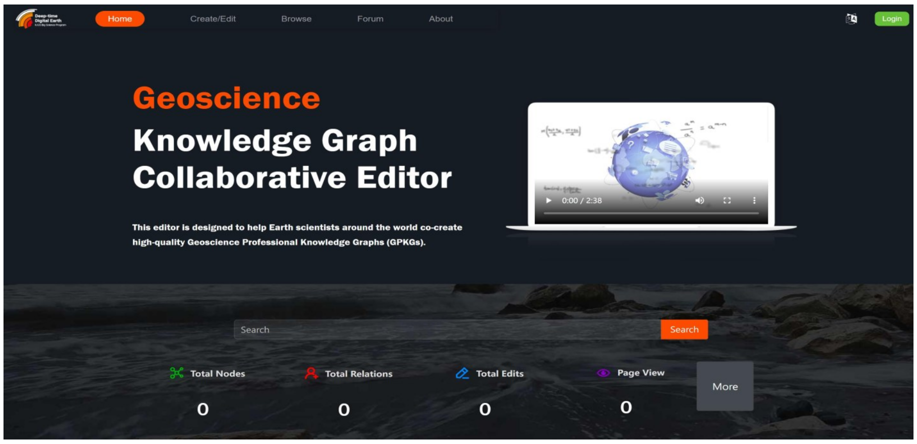

Figure 4. Interface of Geoscience Knowledge Graph Collaborative Editor of DDE (http://editor.deep-time.org/).

DDE KNOWLEDGE PERSPECTIVE

While the current DDE knowledge graph is constructed by domain experts, it remains essential to determine whether the definitions and relationships of these concepts can be established as international standards, thus benefiting a broader community of geoscientists. To enhance the quality and utility of the DDE knowledge graph and ensure its global adoption, two key strategies have been identified for implementation. Firstly, the involvement of more international scholars will be encouraged in the construction and peer review of DDE ontologies. The official societies representing different subfields within geoscience will also be engaged in the construction of the ontologies. Secondly, the GeoOpenKG graph community will be strengthened to facilitate global users' access, editing, and utilization of the graphs, as well as to encourage their feedback on DDE knowledge graph.

Only simple semantic (e.g., hierarchical, causal, constitutive) relationships are built in the current development of DDE ontologies. There is lack of constraints on computational relationships and relationship properties, which directly hinders the inferential capabilities of geoscience knowledge (Zhu et al., 2023c). Therefore, in the future development of geoscience ontologies, domain experts need to express quantitative relationships between concepts and relationship types to enable complex reasoning capabilities within the knowledge graph.

Knowledge graph can be applied to integrate various concepts from databases as mentioned above. Hence, it is essential to enhance the integration of existing concept models from DDE database in the process of constructing domain ontologies. This integration facilitates mapping of database concepts onto a unified ontology schema, thereby facilitating datum integration. As the development of DDE databases continues, it becomes possible to revise, supplement, and standardize closely associated terminologies related to geoscience data. This approach should be a priority for future endeavors in DDE.

The DDE knowledge graph aims to encompass the whole geoscience knowledge and reconstruct its evolution to facilitate the sharing of geoscience knowledge. This requires not only including current concepts but also incorporating outdated ones, such as the term of geosyncline. However, the current expert-built ontologies lack consideration for outdated concepts. Therefore, future DDE knowledge graph construction should adopt a broader bottom-up approach, mining entities and concepts from literature across different time periods, expanding the existing ontology, and assigning a temporal attribute to each concept. This approach will improve understanding of variations in conceptual frameworks among scientists, and facilitate the reconstruction of knowledge evolution in Earth sciences.

ACKNOWLEDGMENTS

The work was financially supported by the National Natural Science Foundation of China (Nos. 42050102, 42050101). This paper contributes to the IUGS "Deep-time Digital Earth" Big Science Program. The final publication is available at Springer via https://doi.org/10.1007/s12583-023-1930-1.

Conflict of Interest

The authors declare that they have no conflict of interest.

REFERENCES CITED

Abel, M., Goldberg, K., De Ros, L. F., 2012. Ontology-Based Rock Description and Interpretation. In: Perrin, M., Rainaud, J. -F., eds., Shared Earth Modeling: Knowledge Driven Solutions for Building and Managing Subsurface 3D Geological Models. Editions Technips, Paris, 261–278

Aggarwal, C., 2021. Artificial Intelligence: A Textbook, Springer, Berlin

Aidan, H., Eva, B., Michael, C., et al., 2022. Knowledge Graphs. *ACM Computing Surveys*, 54(4): 71, https://doi.org/10.1145/3447772

Amit, S., 2012. Introducing the Knowledge Graph. Official Blog of Google

Buttigieg, P. L., Mungall, C., Rueda, C., et al., 2018. SWEET Ontology Suite v3.0.0: Development, Alignments and Use Cases. Abstract from American Association of Geographers 2018, New Orleans

Chen, Q. Z., Yao, H., Li, S. W., et al., 2023. Fact-Condition Statements and Super Relation Extraction for Geothermic Knowledge Graphs Construction. *Geoscience Frontiers*, 14(5): 101412. https://doi.org/10.1016/j.gsf.2022.101412

Gene Ontology Consortium, 2019. The Gene Ontology Resource: 20 Years and still Going Strong. *Nucleic Acids Research*, 47(D1): D330–D338. https://doi.org/10.1093/nar/gky1055

Khider, D., Emile-Geay, J., McKay, N. P., et al., 2019. PaCTS 1.0: A Crowdsourced Reporting Standard for Paleoclimate Data. *Paleoceanography and Paleoclimatology*, 34(10): 1570–1596. https://doi.org/10.1029/2019pa003632

Lexical Systems Group, 2018. The SPECIALIST Lexicon and NLP Tools. https://lhncbc.nlm.nih.gov/LSG

Ma, C., Kale, A. S., Zhang, J. Y., et al., 2023. A Knowledge Graph and Service for Regional Geologic Time Standards. *Geoscience Frontiers*, 14(5): 101453. https://doi.org/10.1016/j.gsf.2022.101453

Ma, K., Tian, M., Tan, Y.J., et al., 2023. Ontology-Based BERT Model for Automated Information Extraction from Geological Hazard Reports. *Journal of Earth Science*, 34(5): 1390–1405. https://doi.org/10.1007/s12583-022-1724-z

Ma, X. G., Fox, P., 2013. Recent Progress on Geologic Time Ontologies and Considerations for Future Works. *Earth Science Informatics*, 6(1): 31–46. https://doi.org/10.1007/s12145-013-0110-x

Ma, X., Mookerjee, M., Hsu, L., Hills, D., 2023. Recent Advancement in Geoinformatics and Data Science. Geological Society of America, Boulder. https://doi.org/10.1130/spe558

Parsons, M. A., Duerr, R., Godøy, Ø., 2023. The Evolution of a Geoscience Standard: An Instructive Tale of Science Keyword Development and Adoption. *Geoscience Frontiers*, 14(5): 101400. https://doi.org/10.1016/j.gsf.2022.101400

Qiu, Q. J., Huang, Z., Xu, D. X., et al., 2023a. Integrating NLP and Ontology Matching into a Unified System for Automated Information Extraction from Geological Hazard Reports. *Journal of Earth Science*, 34(5): 1433–1446. https://doi.org/10.1007/s12583-022-1716-z

Qiu, Q. J., Wang, B., Ma, K., et al., 2023b. A Practical Approach to Constructing a Geological Knowledge Graph: A Case Study of Mineral Exploration Data. *Journal of Earth Science*, 34(5): 1374–1389. https://doi.org/10.1007/s12583-023-1809-3

Shi, S. Z., Lü, H. R., Dong, S. C., et al., 2020. An Editing Platform of Geoscience Knowledge System. *Geological Journal of China Universities*, 26(4): 384–394 (in Chinese with English Abstract)

Simons, B., Boisvert, E., Brodaric, B., et al., 2006. GeoSciML: Enabling the Exchange of Geological Map Data. *ASEG Extended Abstracts*, 2006 (1): 1–4. https://doi.org/10.1071/aseg2006ab162

Tang, X. M., Feng, Z. Q., Xiao, Y. T., et al., 2023. Construction and Application of an Ontology-Based Domain-Specific Knowledge Graph for Petroleum Exploration and Development. *Geoscience Frontiers*, 14 (5): 101426. https://doi.org/10.1016/j.gsf.2022.101426

Wang, C. S., Hazen, R. M., Cheng, Q. M., et al., 2021. The Deep-Time Digital Earth Program: Data-Driven Discovery in Geosciences. *National Science Review*, 8(9): nwab027. https://doi.org/10.1093/nsr/nwab027

Wang, H., Zhong, H. T., Chen, A. Q., et al., 2023. A Knowledge Graph for Standard Carbonate Microfacies and Its Application in the Automatical Reconstruction of the Relative Sea-Level Curve. *Geoscience Frontiers*, 14(5): 101535. https://doi.org/10.1016/j.gsf.2023.101535

Wang, R. J., Yan, Y. C., Wang, J. L., et al., 2018. AceKG: A Large-Scale Knowledge Graph for Academic Data Mining. Proceedings of the 27th ACM International Conference on Information and Knowledge Management. October 22–26, 2018, Torino, Italy. ACM, New York. https://doi.org/10.1145/3269206.3269252

Wang, S., Zhu, Y. Q., Qi, Y. M., et al., 2023. A Unified Framework of Temporal Information Expression in Geosciences Knowledge System. *Geoscience Frontiers*, 14(5): 101465. https://doi.org/10.1016/j.gsf.2022.101465

Xi, J. L., Wu, J., Wu, M. B., 2023. Design and Construction of Lightweight Domain Ontology of Tectonic Geomorphology. *Journal of Earth Science*, 34(5): 1350–1357. https://doi.org/10.1007/s12583-022-1779-x

Xu, H. Q., Zhao, Y. Y., Huang, H., et al., 2023. A Comprehensive Construction of the Domain Ontology for Stratigraphy. *Geoscience Frontiers*, 14(5): 101461. https://doi.org/10.1016/j.gsf.2022.101461

Xu, Y. W., Hu, X. M., Han, Z., 2023. Carbonate Ontology and Its Application for Integrating Microfacies Data. *Journal of Earth Science*, 34(5): 1328–1338. https://doi.org/10.1007/s12583-023-1808-4

Yu, C. M., Zhang, L. M., Hou, M. C., et al., 2023. Climate Paleogeography Knowledge Graph and Deep Time Paleoclimate Classifications. *Geoscience Frontiers*, 14(5): 101450. https://doi.org/10.1016/j.gsf.2022.101450

Zhang, L., Hou, M. C., Chen, A. Q., et al., 2023. Construction of a Fluvial Facies Knowledge Graph and Its Application in Sedimentary Facies Identification. *Geoscience Frontiers*, 14(2): 101521. https://doi.org/10.1016/j.gsf.2022.101521

Zhang, L. N., Hou, Z. S., Shen, B. H., et al., 2023. Paleobiogeographic Knowledge Graph: An Ongoing Work with Fundamental Support for future Research. *Journal of Earth Science*, 34(5): 1339–1349. https://doi.org/10.1007/s12583-023-1845-z

Zhang, X. B., Li, H., Liu, Q., et al., 2023. A New Machine-Learning Extracting Approach to Construct a Knowledge Base: A Case Study on Global Stromatolites over Geological Time. *Journal of Earth Science*, 34(5): 1358–1373. https://doi.org/10.1007/s12583-022-1801-3

Zhou, C., Wang, H., Wang, C., et al., 2021. Prospects for the Research on Geoscience Knowledge Graph in the Big Data Era. *Science China: Earth Sci*., 51(7): 1070–1079 (in Chinese with English Abstract)

Zhu, Y. Q., Dai, X. L., Yang, J., et al., 2023a. One-Stop Sharing and Service System for Geoscience Knowledge Graph. *Geological Journal of China Universities*, 29(3): 325–336 (in Chinese with English Abstract)

Zhu, Y. Q., Sun, K., Hu, X. M., et al., 2023b. Research and Practice on the Framework for the Construction, Sharing, and Application of Large-Scale Geoscience Knowledge Graphs. *Journal of Geo-information Science*, 25(6): 1215–1227 (in Chinese with English Abstract)

Zhu, Y. Q., Sun, K., Li, W. R., et al., 2023c. Comparative Analysis and Enlightenment of Geoscience Knowledge Graphs: A Perspective of Construction Methods and Contents. *Geological Journal of China Universities*, 29(3): 382–394 (in Chinese with English Abstract)

(Responsible Editor: Zhong-Qiang Chen)