Geoscience Frontiers 14 (2023) 101592

Contents lists available at ScienceDirect

Geoscience Frontiers

journal homepage: www.elsevier.com/locate/gsf

Editorial

# The geoscience knowledge system, ontology and knowledge graph for data-driven discovery: Preface

Earth science data have shown rapid growth since the 21st century with the improvement of experimental instruments and testing methods. This provides a basis for revealing the evolutionary history of life, climate, palaeogeography and economic deposits by using big data. However, it is a major challenge to integrate Earth science data for the complexity of the Earth system, the great number of terminologies in Earth science, the diversity of research methods and proxies, and the diversification of data types. To promote the efficient use of big data, the Deep-Time Digital Earth Project (DDE), a big science programme sponsored by the IUGS in 2019 plans to establish a knowledge system and ontology in the field of Earth science, standardize geoscience terminology and unify data standards (Wang et al., 2021; Zhou et al., 2021). To introduce the results of DDE knowledge system construction and the standardization of Earth science data, this special issue of Geoscience Frontiers was organized and it carries eight papers.

The first paper by Ma et al. (2023) established a knowledge graph of the regional geological time scale to harmonize heterogeneous data to facilitate effective and efficient data-driven discovery. This knowledge graph includes the geologic time standards in 17 regions at the Epoch and Age levels. Such a work will help accelerate geologic data integration from multiple sources in data-intensive studies.

The second paper by Wang et al. (2023b) proposed a groundbreaking concept, the Unified Time Framework (UTF), within the geosciences knowledge system. The primary objective of this framework is to enhance the precision and efficiency of calculating and interpreting temporal information across diverse time references. The UTF comprises four key components, namely a root class, general classes, descriptive attributes, and attribute interfaces that connect to external ontologies. By unifying the concept of time in geosciences, the proposed framework will enable the discovery of novel knowledge across various time references facilitating data-driven discovery.

The third paper by Xu et al. (2023) presented the most extensive ontologies ever created for the stratigraphic domain. These ontologies were constructed via federated, crowd intelligence-based collaboration with the aim of implementing state-of-the-art big-data techniques to explore geological history. The comprehensive ontology comprises seven interdependent major branches, including lithostratigraphy, biostratigraphy, chronostratigraphy, chemostratigraphy, magnetostratigraphy, cyclostratigraphy, and sequence stratigraphy. This work serves as a valuable example for constructing domain ontologies in Earth Science and explores the potential applications of semantic search engines based on ontologies for querying references.

The fourth paper by Yu et al. (2023), presented a pioneering hierarchical climate paleogeography knowledge graph. This graph comprises five distinct paleoclimate classifications based on diverse strategies to resolve the issues of inconsistent descriptions and semantic heterogeneity of their climate types. The interrelationships among these climate types in different climate classifications are evaluated through the reconstruction of global climate distributions in the Late Cretaceous. The proposed knowledge graph will assist paleoclimatic correlations and provide a crucial theoretical basis for the application of paleoclimate classifications in deep-time.

The fifth paper by Tang et al. (2023), detailed the construction of an ontology-based domain-specific knowledge graph for petroleum exploration and development using an engineering-based method. The graph is a complex structure, comprising tens of layers, millions of nodes, and over 1000 types of relations. This comprehensive graph is stored in a graph database, and its development is expected to facilitate the creation of intelligent applications and knowledge services for the oil and gas industry.

The sixth paper by Chen et al. (2023), proposed a rule-based multi-input multi-output (R-MIMO) model for information extraction. The authors also constructed a GeothCF dataset, consisting of 1455 fact tuples and 789 condition tuples, to evaluate the performance of their model. By training the R-MIMO model, the authors were able to automatically construct a geothermic knowledge graph with super relations. This geothermic knowledge graph has the potential to facilitate further research in the field of geothermics.

The seventh paper by Parsons et al. (2023) provided an ethnographic examination of how Earth science keywords have evolved and been managed over the past 35 years. The study highlights how semantic approaches have developed over time and provides valuable insights into how standards and associated processes can be sustained and adapted. This work is expected to help formulate various geoscience standards in data-driven research.

The eighth paper by Wang et al. (2023a) presented the development of a standard carbonate microfacies knowledge graph (SMFKG). The authors also created an application that can automatically identify standard carbonate microfacies and reconstruct high-resolution relative sea-level variation curves, which have been successfully applied to the late Ediacaran Dengying Formation in the western margin of the Yangtze terrane. This work is expected to contribute to exploring the Earth system's evolution and predicting future sea-level and climate changes.

Overall, these papers represent the Earth science community's latest efforts toward a comprehensive, machine-readable knowl-

https://doi.org/10.1016/j.gsf.2023.101592

edge system to facilitate the interoperability of big geoscience data. The long-term goal of those efforts is a better data ecosystem to accelerate scientific discoveries in Earth science. We hope this Special Issue can help improve the visibility of the work on ontologies and knowledge graphs in Earth science, as well as lead to more discussion and future work from interested researchers.

**Acknowledgements**

The contributors to this Special Issue and the referees who provided detailed and constructive comments and suggestions are highly acknowledged. Prof. M. Santosh, Prof. Xiaoqiao Wan and Dr. Lily Wang are thanked for the opportunity and editorial assistance. The work was financially supported by the National Natural Science Foundation of China (Grant No. 42050102). The papers in this Special Issue are contributions to the Deep-time Digital Earth (DDE) Big Science Program.

## References

Chen, Q.Z., Yao, H., Li, S.W., Li, X.C., Kang, X.J., Lai, W.W., Kuang, J., 2023. Fact-condition statements and super relation extraction for geothermic knowledge graphs construction. Geosci. Front. 14 (5), 101412. https://doi.org/10.1016/j.gsf.2022.101412.

Ma, C., Kale, A.S., Zhang, J.Y., Ma, X.G., 2023. A knowledge graph and service for regional geologic time standards. Geosci. Front. 14 (5), 101453. https://doi.org/10.1016/j.gsf.2022.101453.

Parsons, M.A., Duerr, R., Godøy, Ø., 2023. The evolution of a geoscience standard: An instructive tale of science keyword development and adoption. Geosci. Front. 14 (5), 101400. https://doi.org/10.1016/j.gsf.2022.101400.

Tang, X.M., Feng, Z.Q., Xiao, Y.T., Wang, M., Ye, T.R., Zhou, Y.J., Meng, J., Zhang, B.S., Zhang, D.W., 2023. Construction and application of an ontology-based domain-specific knowledge graph for petroleum exploration and development. Geosci. Front. 14 (5), 101426. https://doi.org/10.1016/j.gsf.2022.101426.

Wang, C., Hazen, R.M., Cheng, Q., Stephenson, M.H., Zhou, C., Fox, P., She, S.Z., Oberhansli, R., Hou, Z.Q., Ma, X.G., Feng, Z.Q., Fan, J.X., Ma, C., Hu, X.M., Luo, B., Wang, J.L., Schiffries, C.M., 2021. The Deep-Time Digital Earth program: data-driven discovery in geosciences. Natl. Sci. Rev. 8 (9), nwab027. https://doi.org/10.1093/nsr/nwab027.

Wang, H., Zhong, H.T., Chen, Li, K.R., He, H., Qi, Z., Zheng, D.Y., Zhao, H.Y., Hou, M.C., 2023a. A knowledge graph for standard carbonate microfacies and its application in the automatical reconstruction of the relative sea-level curve. Geosci. Front. 14 (5), 101535. https://doi.org/10.1016/j.gsf.2023.101535.

Wang, S., Zhu, Y.Q., Qi, Y.M., Hou, Z.W., Sun, K., Li, W.R., Hu, L., Yang, J., Lv, H.R., 2023b. A unified framework of temporal information expression in geosciences knowledge system. Geosci. Front. 14 (5), 101465. https://doi.org/10.1016/j.gsf.2022.101465.

Xu, H.Q., Zhao, Y.Y., Huang, H., Dong, S.C., Shi, Y.K., Huang, C.J., Wu, H.C., Qian, Z.Q., Fang, Q., Wen, H.G., Su, Z.T., Dai, S., Wang, R.H., Li, C., Sun, C., Fan, J.X., 2023. A comprehensive construction of the domain ontology for stratigraphy. Geosci. Front. 14 (5), 101461. https://doi.org/10.1016/j.gsf.2022.101461.

Yu, C.M., Zhang, L.M., Hou, M.C., Yang, J.H., Zhong, H.T., Wang, C.S., 2023. Climate paleogeography knowledge graph and deep time paleoclimate classifications. Geosci. Front. 14 (5), 101450. https://doi.org/10.1016/j.gsf.2022.101450.

Zhou, C., Wang, H., Wang, C., Hou, Z., Zheng, Z., Shen, S., Cheng, Q., Feng, Z., Wang, X., Lv, H., Fan, J., Hu, X., Hou, M., Zhu, Y., 2021. Prospects for the research on geoscience knowledge graph in the big data era. Sci. China: Earth Sci. 51 (7), 1070–1079. https://doi.org/10.1007/s11430-020-9750-4.

Xiumian Hu *
*School of Earth Sciences and Engineering, Nanjing University, PR China*
* Corresponding author.
*E-mail address*: huxm@nju.edu.cn

Xiaogang Ma
*Department of Computer Science, University of Idaho, USA*

Chao Ma
*Institute of Sedimentary Geology, Chengdu University of Technology, PR China*

Hairong Lv
*Department of Automation, Tsinghua University, PR China*

Available online 27 March 2023