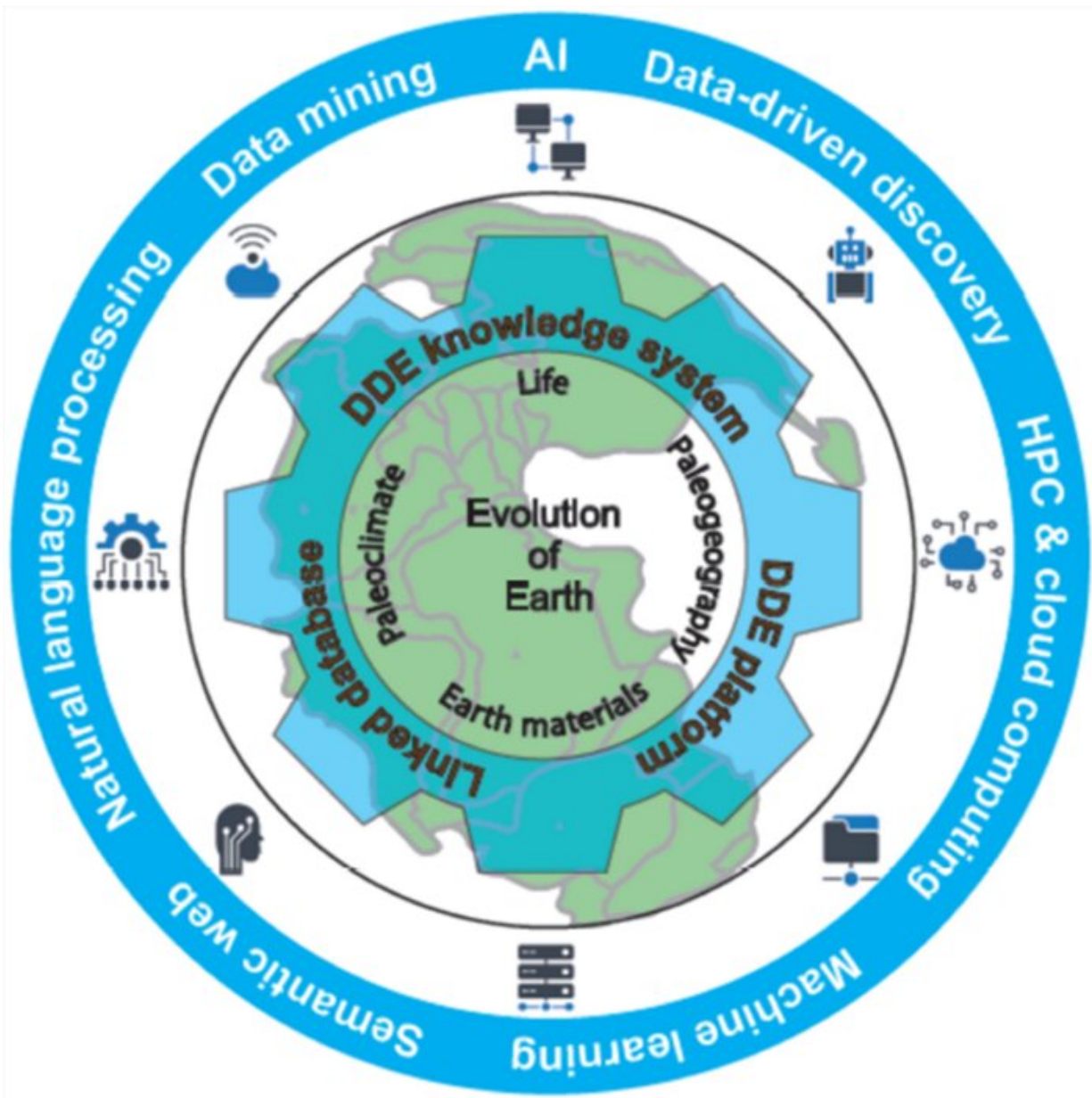# The Deep-time Digital Earth program: Data-driven discovery in geosciences

April 6 2021

DDE aims to harmonize deep-time Earth data based on a knowledge system to investigate the evolution of Earth, including life, Earth materials, geography, and climate. Integrated methods include artificial intelligence (AI), high performance computing (HPC), cloud computing, semantic web, natural language processing, and other methods. Credit: Science China Press

Humans have long explored three big scientific questions: the evolution of the universe, the evolution of Earth, and the evolution of life. Geoscientists have embraced the mission of elucidating the evolution of Earth and life, which are preserved in the information-rich but incomplete geological record that spans more than 4.5 billion years of Earth history. Delving into Earth's deep-time history helps geoscientists decipher mechanisms and rates of Earth's evolution, unravel the rates and mechanisms of climate change, locate natural resources, and envision the future of Earth.

Deductive reasoning and inductive reasoning have been widely employed for studying Earth's history. In contrast to deduction and induction, abduction is derived from accumulation and analysis of large amounts of reliable data, independently of a premise or generalization. Abduction thus has the potential to generate transformative discoveries in science. With the accumulation of enormous volumes of deep-time Earth data, geoscientists are poised to transform research in deep-time Earth science through data-driven abductive discovery.

However, three issues must be resolved to facilitate abductive discovery using deep-time databases. First, many relevant geodata resources are not in compliance with FAIR (findable, accessible, interoperable and reusable) principles for scientific data management and stewardship. Second, concepts and terminologies used in databases are not well defined; thus, the same terms may have different meanings across

databases. Without standardized terminology and definitions of concepts, it is difficult to achieve data interoperability and reusability. Third, databases are highly heterogeneous in terms of geographic regions, spatial and temporal resolution, coverages of geological themes, limitations of data availability, formats, languages and metadata. Due to the complex evolution of Earth and interactions among multiple spheres (e.g., lithosphere, hydrosphere, biosphere and atmosphere) in Earth systems, it is difficult to see the whole picture of Earth's evolution from separated thematic views, each with limited scope.
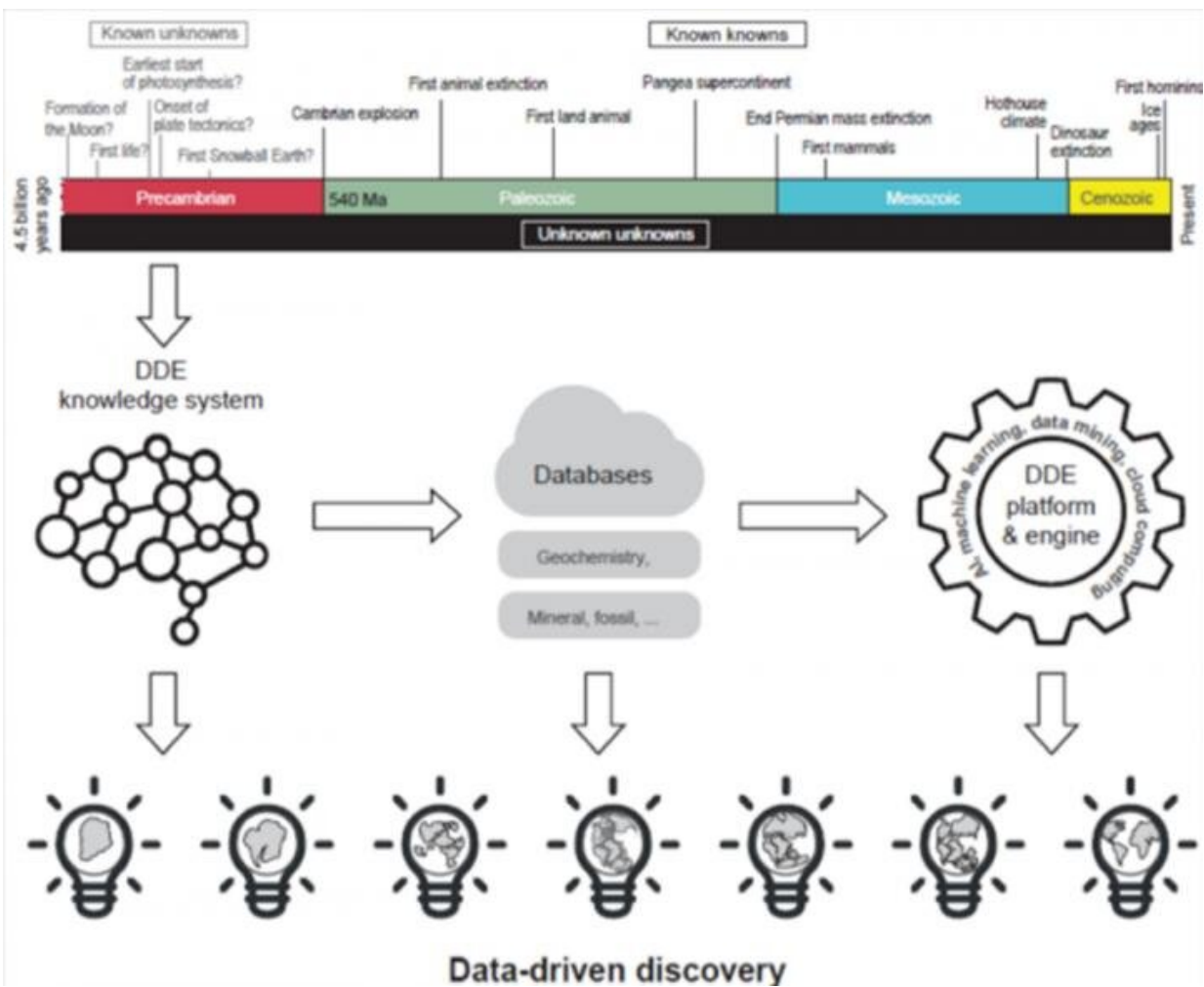
Scientific questions in Earth history can be addressed using the knowns and

unknowns framework: (1) Known knowns. This category, which is relative to the other two, includes widely accepted and broadly understood events in Earth history, although uncertainties still exist. (2) Known unknowns. This category includes events that are widely accepted to have happened but key aspects are poorly understood. In many cases, hypotheses about such events can be tested with additional observations, measurements, or experiments. (3) Unknown unknowns. This category includes events that took place in the Earth's history but have not been discovered. Through its knowledge system and platform, DDE aims to harmonize deep-time Earth data and promote data-driven discovery in these unknowns, especially unknown unknowns in Earth history. Note: the time scale of Precambrian and Phanerozoic are differ in scale. Credit: Science China Press

Big data and artificial intelligence are creating opportunities for resolving these issues. To explore Earth's evolution efficiently and effectively through deep-time big data, we need FAIR, synthetic and comprehensive databases across all fields of deep-time Earth science, couple with tailored computation methods. This goal motivates the Deep-time Digital Earth program (DDE), which is the first "big science program" initiated by the International Union of Geological Sciences (IUGS) and developed in cooperation with national geological surveys, professional associations, academic institutions, and scientists around the world. The main objective of DDE is to facilitate deep-time, data-driven discoveries through international and interdisciplinary collaborations. DDE aims to provide an open platform for linking existing deep-time Earth data and integrating geological data that users can interrogate by specifying time, space, and subject (i.e., a "Geological Google") and for processing data for [knowledge discovery](link) using a knowledge engine (Deep-time Earth Engine) that provides computing power, models, methods, and algorithms (Figure 1).

To achieve its mission and vision, the DDE program has three main

components: program management committees, centers of excellence, and working, platform and task groups. And DDE will build on existing deep-time Earth knowledge systems and develop an open platform (Figure 2). A deep-time Earth knowledge system consists of the basic definitions and relationships among concepts in deep-time Earth, which are necessary for harmonizing deep-time Earth data and developing a knowledge engine for supporting abductive exploration of Earth's evolution. The first step in DDE's research plan is to build on existing deep-time Earth knowledge systems. The second step in DDE's research plan is to build an interoperable deep-time Earth data infrastructure. And the third step in DDE's research plan is to develop a deep-time Earth open platform.

The execution of the DDE program consists of four phases. In Phase 1, DDE establishes an organizational structure with international standards of policy and management. In Phase 2, DDE forms the initial teams and builds on existing deep-time Earth knowledge systems and data standards by collaborating with existing ontology researchers in the geosciences, while working to link and harmonize deep-time Earth databases. In Phase 3, DDE develops tailored algorithms and techniques for environments of cloud computing and supercomputing. In Phase 4, Earth scientists and data scientists collaborate seamlessly on compelling and integrative scientific problems.

As integrative and international ambitions of the DDE program, several challenges were anticipated. However, by creating an open-access data resource that for the first time integrates all aspects of Earth's narrated past, DDE holds the promise of understanding our planet's past, present, and future in new and vivid detail.

More information: Chengshan Wang et al, The Deep-time Digital Earth program: data-driven discovery in geosciences, *National Science Review* (2021). DOI: 10.1093/nsr/nwab027

Provided by Science China Press

Citation: The Deep-time Digital Earth program: Data-driven discovery in geosciences (2021, April 6) retrieved 12 May 2024 from https://phys.org/news/2021-04-deep-time-digital-earth-data-driven-discovery.html