

ON THE CUSP OF A **REVOLUTION**

In geoscience, we run the risk of unfulfilled potential in the world of big data analytics. **Mike Stephenson** and colleagues suggest the Deep Time Digital Earth project could help by providing a global data platform, or 'geological Google'





In the era of big data, the ease of availability of digital data sets and new approaches to data analysis, such as artificial intelligence and machine learning, are revolutionising many scientific fields.

Geoscientific fields are benefitting too, but most successful applications of big data analytics are limited to more standardised data types, such as seismic data. Major advances in geoscience are hampered by the general complexity and heterogeneity of geological data, as well as the reliance on proxies as indirect measurements. Additionally, the field can suffer from poor data management practices and lack of linkages between established databases.

These data are important. They represent a untapped source of rare information that might harbour some secret to our planet's history or future. By successfully standardising and linking up disparate data sets, we could apply big data analytical approaches and potentially make revolutionary discoveries. The Deep Time Digital Earth (DDE) programme aims to do just that.

Deep Time Digital Earth

A few wintry days in February 2019 at the Fragrant Hills Hotel Beijing saw a gathering of 70 top geologists and geoscience data specialists. The hotel is within a spectacular park with buildings that hosted rulers of the Yuan, Ming and Qing dynasties between the 12th and 18th centuries and contains the Shuangqing Villa, once the residence of Mao Zedong. But this gathering of geoscientists and representatives of many of the world's foremost international geoscience associations came to discuss how to bring geoscience data into the 21st century.

The new initiative of the International Union of Geological Sciences (IUGS) known as the Deep Time Digital Earth (DDE) programme aims to link geological databases and

make them accessible seamlessly from one portal—rather like a geoscience search engine or 'geological Google'.

The time for an ambitious programme of this type has never been better. Sensor data from the geosciences are being provided in ever-larger quantities, sensors for geological applications are becoming more available and cheaper, and computing power and visualization is well able to cope with floods of geoscience data.

'Long-tail' data

DDE's aim, though, is not primarily to collect sensor data, but to focus on so-called 'long-tail' data—the difficult-to-get-at data that sit in institutes, libraries and on the computers of individual scientists. Informatics specialists contrast such data with the smaller number of large, more accessible data sets associated with sensors. The name 'long-tail' derives from graphs drawn of the size of data sets against their number: there are relatively few large datasets and many smaller ones (Fig. 1). Geological science has more long-tail data than sciences like physics or meteorology, probably because historically it has been less associated with big science infrastructure and sensors. Much long-tail geological data are concerned with 'deep time' or pre-Quaternary time.

The fact that long-tail deep-time data are difficult to access holds back progress in geological science. The low discoverability and heterogeneity of such data make them difficult to bring together to gain the benefits of machine learning and artificial intelligence, and to see the data in full detail. Modern sensor data are used to measure and model present-day Earth processes, but it is deep-time data that we need to understand past processes and events. Through DDE, these data will be available in easy-to-use hubs. Data brought together in new ways will afford insights into the distribution and value of Earth's resources and materials, as well as its hazards, and ►

► may provide novel glimpses into Earth’s geological past and future.

Linking databases will be a big part of DDE. An example is the integration of deep-time data for mapping clusters of porphyry copper mineral deposits (PCDs). It’s thought that deep-time plate motion, as well as crustal and slab subduction help control the distribution of PCDs. Some of these processes are described in established databases and models, for example, in relation to crustal thickness in the Crust 1.0 Model, the Global Plate Reconstruction Model,

the Slab 2.0 Model and the global U-Pb database. Coupling between these models and databases is possible, but it takes a long time and significant computing skills. The aim of DDE will be to do the hard work in advance, by linking georeferenced databases and models of this type together so that they can be used more efficiently. Data will still reside with the originating institution or individual, so ownership will not change—it’s just that the links between databases and models will be better (Fig. 2).

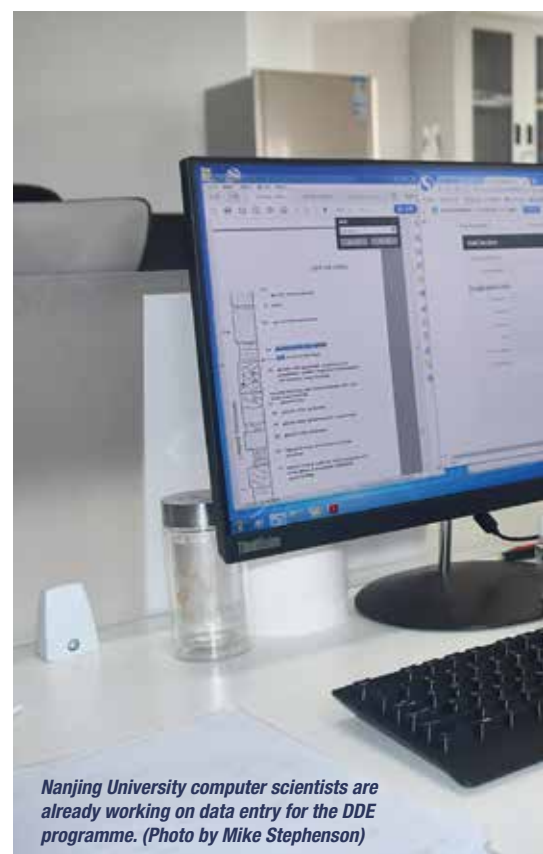
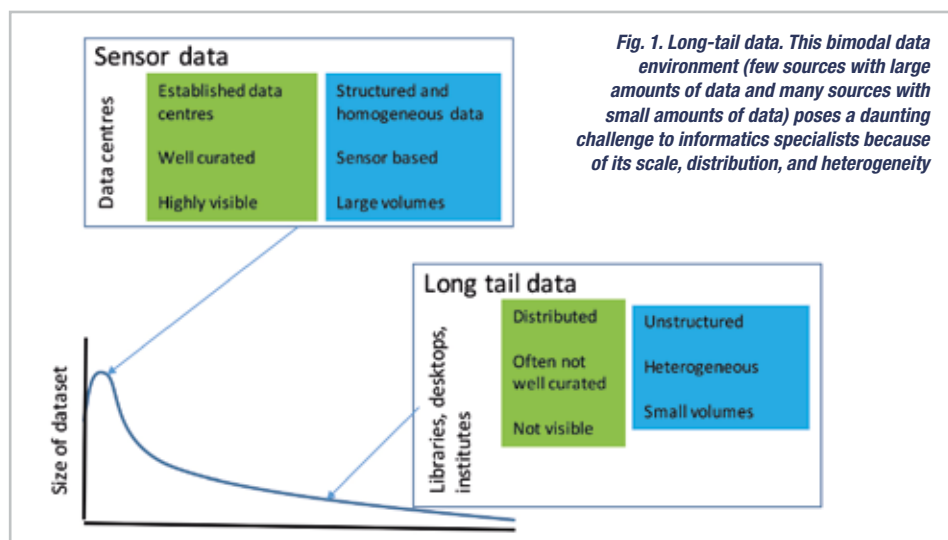
Another example of how DDE will work concerns the evolutionary history of the biosphere. Previous analyses of long-term palaeobiodiversity change were mostly at a resolution of about 10 million years, which is too coarse to reveal the fine details of any changes. Databases linked through DDE could provide high-resolution (10 to 100 thousand year) diversity patterns.

Yet other DDE linkages will aim to address the sustainable development goals. For example, databases and models could be coupled to provide better understanding of African groundwater storage, thereby helping countries vulnerable to climate change.

Vision of the IUGS

The DDE is closely consistent with the vision of the IUGS, which is to promote development of the Earth sciences through the support of broad-based scientific studies relevant to the entire Earth system.

DDE is a truly global initiative. It brings together a unique range of founding members including the International Commission on Stratigraphy, the International Palaeontological Association, the International Association of



Sedimentologists, the Society for Sedimentary Geology, the American Association of Petroleum Geologists, and the International Association for Mathematical Geosciences. Major geological surveys, institutes and commissions are also involved, including the China Geological Survey, the British Geological Survey, the All Russian Geological Institute, the Commission for Geological Map of the World and the Commission on the

Management and Application of Geoscience Information.

DDE will operate under the full Findable, Accessible, Interoperable, and Re-usable (FAIR) data concept. It will link to the desktop systems of geoscientists all over the world, as well as to students and teachers in classrooms and on the internet. These institutions are coming together at a time when informatics and computing are evolving fast, but where a wider

range of geoscience data were not available until now. In this way, DDE may help to solve some of the biggest geoscience questions that remain. The DDE programme is being developed right now, ready for its launch at the International Geological Congress in March next year where geoscientists from all over the world will be able to hear about activities and get involved.

The DDE vision has stimulated large amounts of funding, including \$75 million from the Government of China to build a DDE-dedicated centre of excellence at Suzhou near Shanghai, with access to one of the world's fastest supercomputers, the Shenwei TaihuLight. In the United States, plans are developing for a DDE centre of excellence looking at geological resources. In Europe, it is hoped that the DDE can be linked with the 'OneGeology' concept (an international initiative of geological surveys working together to make geoscience data web-accessible worldwide).

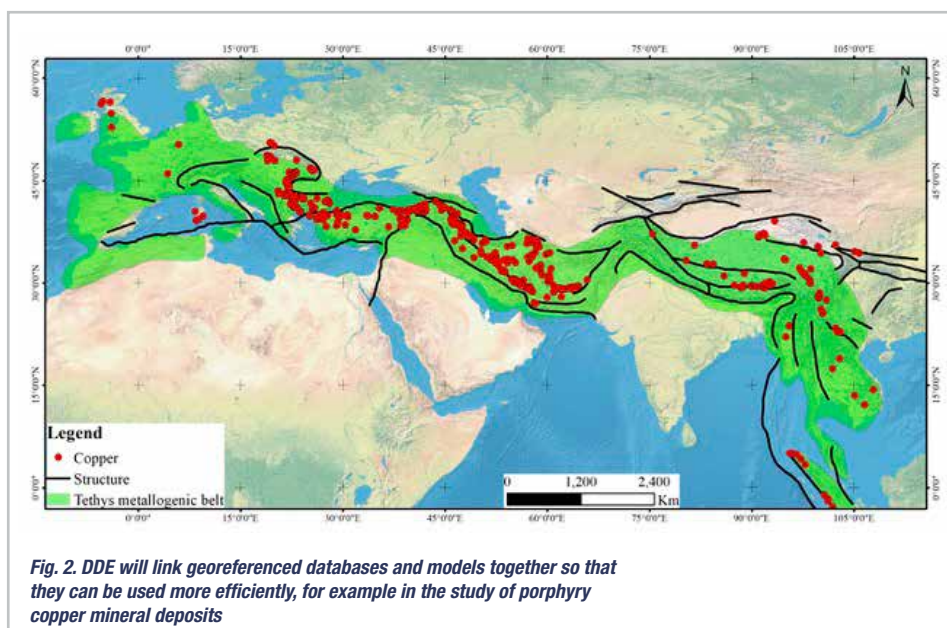


Fig. 2. DDE will link georeferenced databases and models together so that they can be used more efficiently, for example in the study of porphyry copper mineral deposits



Building bridges

DDE will enable the building of bridges between data islands and allow data to be interrogated using modern tools, thereby tackling some of the most important and pressing questions of our time. However, a big part of DDE will also be building communities of like-minded scientists. It could be said that geology has lagged behind other physical sciences in capitalizing on big data—but with DDE, geoscience will catch up. Perhaps some big geological discoveries still lie ahead of us?

Prof Mike Stephenson is President of the Governing Council of DDE, Executive Chief Scientist at the British Geological Survey and Visiting Professor at Nanjing University. Email mhste@bgs.ac.uk

Prof Qiuming Cheng is President of the International Union of Geological Sciences

Prof Chengshan Wang is a member of the Chinese Academy of Sciences and Professor at the China University of Geosciences

Prof Junxuan Fan, is a Professor at the Nanjing University

Prof Roland Oberhänsli is Past-President of the IUGS.

Further information on the Deep Time Digital Earth programme can be found at <http://www.ddeworld.org>