

by Michael H. Stephenson^{1*}, Qiuming Cheng², Chengshan Wang³, Junxuan Fan⁴, and Roland Oberhänsli⁵

Progress towards the establishment of the IUGS Deep-time Digital Earth (DDE) programme

¹ British Geological Survey, Keyworth, Nottingham, NG12 5GG, UK; *Corresponding author, *E-mail*: mhste@bgs.ac.uk

² School of the Earth Science and Resources, China University of Geosciences, Beijing 100083, China

³ State Key Laboratory of Biogeology and Environmental Geology, China University of Geosciences, Beijing 100083, China

⁴ School of Earth Sciences and Engineering, Nanjing University, Nanjing 210023, China

⁵ Potsdam University, Institute of Geosciences, Karl Liebknechtstrasse 24, 14476 Potsdam, Germany

(Received: February 1, 2020; Revised accepted: March 31, 2020)

<https://doi.org/10.18814/epiugss/2020/020057>

The Deep-time Digital Earth (DDE) programme of the International Union of Geological Sciences (IUGS) has been developed to address the formidable challenge of so called ‘long tail’ data in the geosciences - the unstructured and inherently heterogeneous geoscience data that resides in institutions, universities and on individual geoscientists’ computers. DDE’s vision is to transform Earth science by connecting and harmonising long tail deep-time data ‘islands’ to support broad-based scientific studies relevant to the entire Earth system. The results of these and other studies will help us understand Earth’s natural environment and help in the wise use of natural resources for the prosperity of nations and the quality of human life. This harmonisation is now possible through the digital revolution, but new protocols, platforms and programs are needed to secure compatible and interoperable databases, so that the vast amounts of existing (and new) deep-time geoscience data can be linked. Since the first DDE meeting in January 2019, great progress has been made in defining statutes and byelaws, governance structures and preliminary informatics and scientific aims.

Introduction

Geoscience as a scientific discipline stands at a point where it can benefit from the great advances that data could provide: in analysis, computation and visualization. It could benefit in the way that atmospheric and theoretical physics have benefitted, allowing access to big data techniques to solve some of the big physical science questions of the Earth.

Over the last twenty or thirty years access has improved to large databases hosted at established data centres, such as the U.S. Geological Survey (e.g., Eros data centre, <http://eros.usgs.gov/>), NASA (e.g., Atmospheric Science Data Centre, ASDC, at NASA Langley Research Centre (<http://eosweb.larc.nasa.gov/>), NOAA (e.g., National Geophysical

Data Centre, NGDC, <http://www.ngdc.noaa.gov/>), UCAR (e.g., Data for Atmospheric Research, <http://rda.ucar.edu/>), EPOS (<https://www.epos-ip.org/>), and others.

Although availability of these sensor-based data is crucial for the research goals of geoscientists, a vast amount of data generated by individual scientists and institutions is still not accessible; even its existence may not be known. The situation has a parallel in medical and health science where data have become more widely interoperable and accessible. Many of these advances in medical science were made in the post-war period in the United States where computing began to be used on hitherto unavailable data to solve problems in health care and medicine. This effort of gathering data, improving interoperability and accessibility, has meant that, as in some of the physical sciences, new techniques of artificial intelligence and machine learning have brought enormous advances in diagnosis and epidemiology, for example in the successful machine recognition of cancerous cells in microscope slides.

Geoscience data have some of the characteristics of health and medical data from the early Twentieth Century in that historical observations, or legacy data that are typically found in what is referred to as the “long tail” (e.g., Sinha et al., 2013; Fig. 1) are needed to advance science. Making these data available on demand is one of the highest priorities to examine the Earth system through time.

Typically, sensor data are used to measure, evaluate, and model present-day Earth processes, but are unable to capture past processes and events. Development and modelling of more complex, temporal views of the Earth requires many data types, instruments, techniques, and computational services.

Two fundamentally different types of data management are required: those generated by the long tail of science, and those generated by sensor technologies. The latter reside in well-designed and curated data centres, with the only significant challenge lying in evolving technologies that may make the earlier data unusable or inaccessible. In contrast, the long tail of geoscience is more unstructured and inherently more heterogeneous. DDE’s vision is to transform Earth science by connecting and harmonising long tail deep-time data ‘islands’ to support broad-based scientific studies relevant to the entire Earth system.

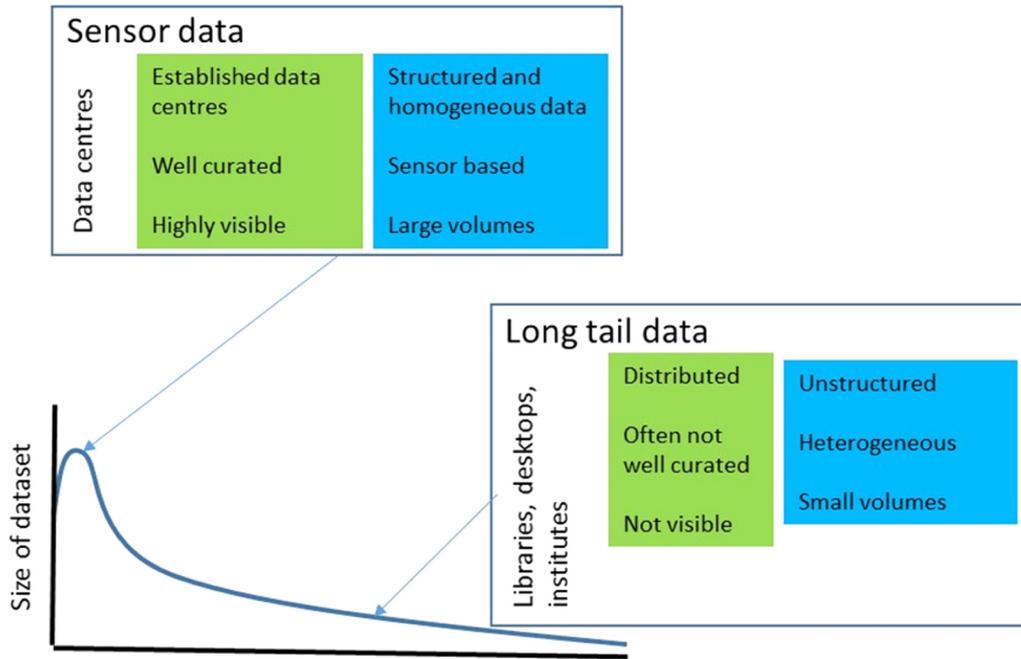


Figure 1. Two types of geoscience data.

This harmonisation is now possible through the digital revolution, but new protocols, platforms and programs are needed to secure compatible and interoperable databases, so that the vast amounts of existing (and new) deep-time geoscience data can be linked. Big data analytics, internet cloud computing, data mining, machine learning and artificial intelligence, alongside traditional geological research will lead to innovation in understanding the Earth’s evolution and applications including the United Nations Sustainable Development Goals.

The philosophical approach of DDE will be to provide the ‘wiring’ that will connect disparate and distributed deep-time database ‘islands’ (Fig. 2).

The DDE program will address ten grand challenges, which will provide better understanding of major global issues such as: the origin and evolution of life on earth, the dynamics of the earth-life systems, prediction

of global change, and sustainability for the future earth. The initial challenge for DDE will be to secure interoperability of databases, as well as accessibility. This better organized data will transform Earth science. From this transformative stage will come a number of innovations including a high-resolution integrative timeline for earth history, machine reading technologies for discovery and utilization of digital data and published documents, digitization and comparison of maps and graphic data, feature recognition in real-time fieldwork, and evaluation and spatiotemporal association of big data.

DDE and the International Union of Geological Sciences

The vision for DDE was discussed at an introductory meeting in

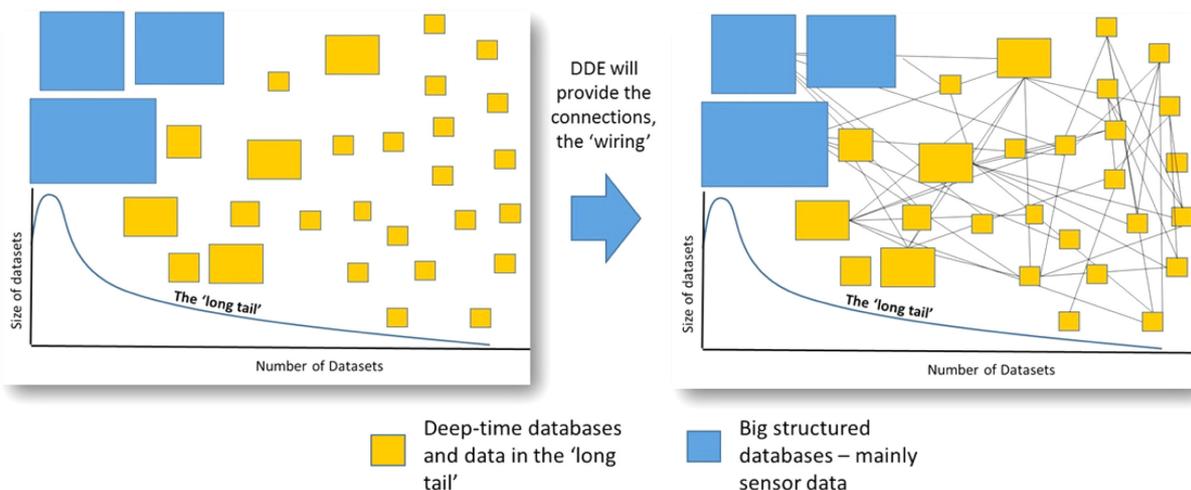


Figure 2. New protocols, platforms and programs are needed to secure compatible and interoperable databases, so that the vast amounts of existing (and new) deep-time geoscience data can be linked. DDE will provide the wiring to connect deep-time data sources together.



Figure 3. Geoscientists from the IUGS and DDE founding members at the Fragrant Hills Hotel, Beijing.

February 2019 at the Fragrant Hills Hotel in Beijing, China, at which gathered seventy top international geologists and geoscience data specialists as well as representatives of the International Union of Geological Sciences (IUGS; Fig. 3; Wang et al., 2019). The meeting was hosted by Prof Qiuming Cheng, President of the IUGS, and the abstracts of the meeting are published in *Acta Geologica Sinica* (Fan et al., 2019).

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. At the Beijing Fragrant Hills meeting, the early development of the vision enabled the support of an almost unique range of founding members including the International Commission on Stratigraphy, the International Association of Palaeontology, 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. Through its founding members, DDE aims to be a truly global initiative. DDE will operate the full FAIR data concept (Findable, Accessible, Interoperable, and Re-usable; Wilkinson et al., 2016) and link to desktop systems for 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 still remain.

As part of the fragrant hills meeting, guests from the international community and IUGS representatives were invited to Suzhou near



Figure 4. An event to launch a DDE-dedicated centre of excellence at Suzhou near Shanghai.

Shanghai to hear about plans for the development of a DDE-dedicated centre of excellence (Fig. 4). The DDE vision has been compelling enough so far to stimulate large amounts of funding including \$75 million to build the centre.

Following these meetings, the steering committee: Qiuming Cheng, President of the International Union of Geological Sciences, Chengshan Wang, Professor at the China University of Geosciences, Junxuan Fan, Professor at Nanjing University, Roland Oberhänsli Past-President of the IUGS and Michael Stephenson, of the British Geological Survey, have met regularly to progress DDE's aims. A meeting in October 2019, again in Beijing, further consolidated the case for DDE and refined its vision and mission, policy on governance and its scientific direction.

Mission, Vision and Governance of DDE

DDE’s mission is to harmonise global Deep-time Digital Earth data, and share global geoscience knowledge. Its ambitious vision is to ‘transform Earth science’.

DDE is an ambitious global initiative that already involves, and will in the future involve, many stakeholders and types of stakeholders, from public institutions like libraries and national research institutions and geological surveys, to educational institutions like universities, to Government departments, and to learned societies and professional associations. This variety of stakeholders will reflect national and cultural differences as well widely different approaches to internal governance and ways of working.

Governance is essentially the structure and processes that exist in and between formal institutions that allows them to work together for a common goal, and to allow the many stakeholders to see how they benefit. Good governance means that all stakeholders buy-in to the initiative, and can see its benefits and feel part of its decision-making. Good governance leads to a well-functioning initiative.

DDE will bring together a very wide range stakeholders and so has a strong internal governance to ensure efficient decision-making and business processes. The conspicuously transparent governance structure will also foster inclusion externally, and ensure legitimacy and encourage trust (Fig. 5). DDE will work with partner organisations towards ensuring diversity and gender balance in all its governance boards and committees, and in forming the collaborations and research teams that will deliver DDE science and informatics solutions.

The DDE Board will be the highest level governance of the DDE programme. The aim is for ‘light touch’ scrutiny of DDE programmes including review of performance against a limited number of key performance indicators (KPIs) set by the DDE Executive Committee. The Board will not take executive decisions nor be involved in the day-to-

day management or running of DDE.

The Governing Council (GC) controls all day-to-day activities of DDE, and it will consist of one delegate nominated by each of the founding members of DDE, and voting in the Governing Council will be on the basis of one vote per founding member. The Executive Committee will carry out the administration of DDE, with the assistance of the DDE Secretariat. The Science Committee evaluates proposals to DDE and ranks them for further approval by the GC. DDE Working groups focus on knowledge systems of disciplines and carry out the scientific activities of DDE. Task groups focus on exploring databases through projects and/or meetings arranged if possible in collaboration with the founding members, IUGS, other non-governmental bodies, and inter-governmental bodies.

Benefits of DDE

The benefits of DDE can be divided into two. The first and most important is in relation to its stated aim to organise data – making data more accessible and databases interoperable will make tackling science questions easier, for example evolution of minerals and continental cycles (e.g., Fox, 2019). The second involves the formulation and strategic aim to tackle geoscience questions that follow from and are amenable to the liberation of data (e.g., Nash, 2019; Montanez et al., 2019). DDE will enable bridges between data islands to be built and for data to be interrogated using modern tools tackling some of the most important and pressing questions of our time. Some of the provisional scientific questions and aims include:

1. Evolution of Life
 - a. How did life and biodiversity originate and evolve?
 - b. How did ecosystems respond to extreme climatic conditions?
 - c. How to build a diversity pattern for the terrestrial ecosystem?

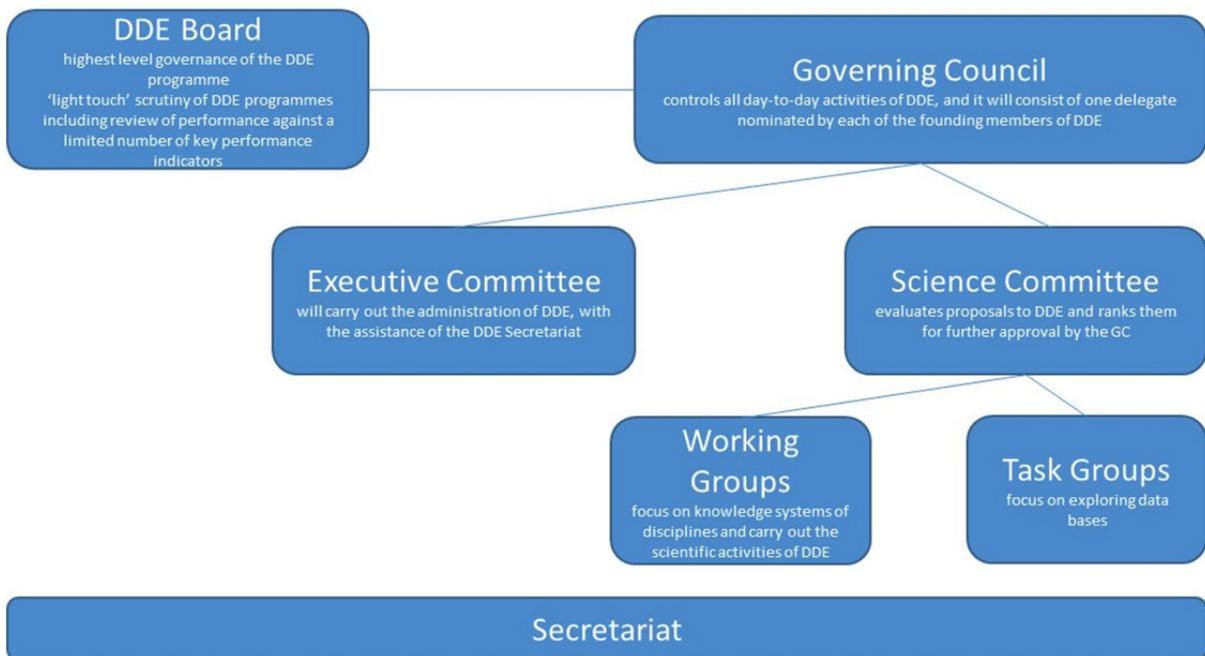


Figure 5. Governance scheme of DDE.

- d. Understanding the cycles and evolution of bio-elements in the ocean
2. Evolution of materials
 - a. How did sedimentary materials evolve and cycle?
 - b. Four dimensional (4D) architecture and evolution of deep-earth materials and dynamics
3. Mineral evolution beyond 4D
 - a. Evolution of climate
 - b. Reconstructing Earth's climate and atmosphere history
4. Chemical evolution of Earth's oceans
 - a. Evolution of geography
 - b. Global sea-level change through deep time
 - c. Quantifying plate tectonics and deformation in four dimensions
 - d. How did the geomagnetic field co-evolve with the deep time Earth system?

DDE will also seek to solve many applied geological problems in the area of energy and resources, working with geological surveys, companies and governments to improve energy and resource security, tackle modern day climate change, and to contribute to the UN's Sustainable Development Goals (SDGs; Stephenson, 2019, 2020). This will be achieved by linking databases together to answer applied geoscience questions, recognising the spatially (and temporally) explicit nature of the methods which geostatistics provides for the appropriate analysis of earth science properties, which vary in space and time (McKinley and Atkinson, 2020)

Examples of questions might include the integration of deep-time data for mapping clusters of porphyry copper mineral deposits (PCDs; e.g. Cheng, 2019). It is thought that deep time plate motion and crustal and slab subduction all have roles in the distribution of PCDs and some of these processes are described in established databases and models, for example in relation to crust thickness (Crust 1.0 Model), the Global Plate Reconstruction Model, the Slab 2.0 model and the global U-Pb database. Coupling of these models and databases is possible, but takes a long time and significant computing skills. Data analysis tools of linked databases and models can be used to gain insights into PCDs that are not possible through simple analysis of single or pairs of databases. DDE will link georeferenced databases and models of this type together so that they can be used more efficiently.

In Africa groundwater is the major source of drinking water and its use for irrigation will probably increase as Africa's economy and population grow. Despite its importance, there is a dearth of detailed quantitative information on groundwater in Africa. Often groundwater storage is not factored in to assessments of freshwater availability. Clearly a prediction involving less runoff in climate change forecasts has serious implications for the population that depend on water. How much of the deficit can be compensated for by, for example, groundwater? Are the aquifers large enough in the area and will water wells deliver the deficit? Will recharge in the area maintain the sustainability of the aquifer and the water wells or will it be affected by climate change in the area? Similarly what would the effects be of a large increase in runoff? What measures might have to be put in place to deal with flooding? Much of the research needed to answer these questions locally has still to be done. Linked databases in DDE, including African groundwater and aquifer data, recharge data, meteorological data, sediment flux, subcrop geology, basin subsidence, sequence stratigraphy, compaction, geomechan-

ics and tectonics data could lead to more accurate models for African sedimentary basins, groundwater storage including aquifer properties and structure, underpinning sustainable development in some of the poorest countries, and the countries most vulnerable to climate change (e.g., Stephenson, 2019).

Integrated georeferenced databases for sediment flux, subcrop geology, basin subsidence, sequence stratigraphy, compaction, geomechanics and tectonics could lead to simulation of basin development ('digital twin' basins) and to 'forward stratigraphic modelling' to provide information on the distribution of facies, diagenesis and faulting in important resource basins. Across Europe, Palaeozoic, Mesozoic and Tertiary basins particularly provide access to resources such as groundwater, minerals, materials for infrastructure and energy resources and storage, for example the Chalk aquifer that extends across northern Europe supplies groundwater, and the Permo-Triassic salt provinces of Europe, North Africa and the Atlantic Margins that provide large-scale underground storage opportunities (e.g., Stephenson, 2019).

Heterogeneities (facies, diagenesis, faults) govern the utility of these lithological units and basins for future use. For example, understanding of the heterogeneity of Cretaceous chalk lithofacies is vital for gas storage, hydrocarbon potential, and sustainability as an aquifer. The latter is important over wide areas of northern Europe where the chalk is an important aquifer and will be more important as climate change proceeds. The northern European Permo-Triassic salt provides specialist gas storage for example for compressed air energy storage (CAES) for energy intermittency and hydrogen storage for a possible hydrogen economy with decarbonised domestic heat and a hydrogen transport fleet (e.g., Stephenson, 2018, 2019). Linked databases in DDE will enable 'digital basins' to be developed to simulate the growth of basins and their heterogeneities and the management of basin usage (e.g., pressure management, groundwater control and management).

Conclusions

DDE is an emerging movement to tackle the challenge of so called 'long tail' data in the geosciences, the unstructured and inherently heterogeneous geoscience data that resides in institutions, universities and on individual geoscientists' computers. DDE's vision is to transform Earth science by connecting and harmonising long tail deep-time data 'islands' to support broad-based scientific studies relevant to the entire Earth system.

Acknowledgements

Mike Stephenson publishes with the permission of the Director of the British Geological Survey.

References

- Cheng, Q., 2019, Integration of Deep-time Digital Data for Mapping Clusters of Porphyry Copper Mineral Deposits. *Acta Geologica Sinica*, v. 93, S3, pp. 8–10. doi:org/10.1111/1755-6724.14228
- Fox, P., 2019, Disruption in Biogeosciences: Conceptual, Methodological, Digital, and Technological. *Acta Geologica Sinica*, v. 93, S3, pp.

17–18. doi:org/10.1111/1755-6724.14231

McKinley, J.M., and Atkinson, P.M., 2020, A Special Issue on the Importance of Geostatistics in the Era of Data Science. *Mathematical Geosciences*, v. 52, pp. 311–315.

Montañez, I.P., Hu, X., Hou, M., Wang, C., and Chen, J., 2019, Sedimentological/Palaeogeographic Data as Fundamental Building Blocks of the DDE Project: Critical Underpinning of Reconstructing Deep-time Earth Systems. *Acta Geologica Sinica*, v. 93, S3, pp. 52–54. doi: org/10.1111/1755-6724.14242

Nash, S.S., 2019, The Role of Big Data and Machine Learning in the Integration and Implementation of Historical, Current, and Continuously Gathered Earth Data. *Acta Geologica Sinica*, v. 93, pp. 56–58. doi: org/10.1111/1755-6724.14242

Sinha, A.K., Thessen, A.E., and Barnes, C.G., 2013, Geoinformatics: Toward an integrative view of Earth as a system. In: Bickford, M.E., (Ed.), *The Web of Geological Sciences: Advances, Impacts, and Interactions*. Geological Society of America Special Paper 500, pp. 591–604. doi:10.1130/2013.2500(19)

Stephenson, M.H., 2018, *Energy and Climate Change: An Introduction to Geological Controls, Interventions and Mitigations*. Elsevier, Amsterdam, 186 p.

Stephenson, M.H., 2019, The Uses and Benefits of Big Data for Geological Surveys. *Acta Geologica Sinica*, v. 93, S3, pp. 64–65. doi: org/10.1111/1755-6724.14247

Stephenson, M.H., in press for 2020, Affordable and clean energy as a sustainable development goal: energy geoscience questions and human capacity. In: Gill, J.C., Smith, M. (Eds.) *Geosciences and the Sustainable Development Goals*. Springer International Publishing, ISBN 978-3-03038814-0.

Wang, C., Shen, S., Zhou, C., Hou, Z., Cheng, Q., Oberhänsli, R., and Stephenson, M., 2019, Decodification of Earth Evolution in Deep-Time. *Acta Geologica Sinica*, v. 93, S3, pp. 1–2.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., and Appleton, G., 2016, The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, v. 3: 160018. doi:10.1038/sdata.2016.18

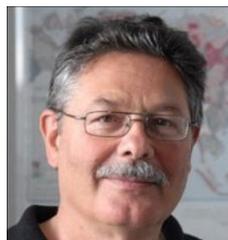
Zhong, Y., Chen, D., Fan, J., Wu, H., Fang, Q., and Shi M., Cyclostratigraphic Calibration of the Upper Ordovician (Sandbian-Katian) Pagoda and Linhsiang Formations in the Yichang Area, South China. *Acta Geologica Sinica*, v. 93, S3, pp. 177–180. doi:org/10.1111/1755-6724.14282



Qiuming Cheng is a Chinese mathematical geoscientist. He is a professor and founding director of the State Key Lab of Geological Processes and Mineral Resources, China University of Geosciences (Beijing). He received the William Christian Krumbein Medal in 2008 from the International Association for Mathematical Geosciences. He was the President of the International Association for Mathematical Geosciences (2012–2016). He is currently the President of the International Union of Geological Sciences (IUGS).



Chengshan Wang is a sedimentologist and Professor at China University of Geosciences (Beijing). He has been elected as Fellow of the Geological Society of America in 2012, and Member of the Chinese Academy of Sciences in 2013. Chengshan currently is the Vice-President of the International Association of Sedimentology (IAS). His research interest and main activities are related to paleogeography and paleoclimate reconstruction. He used to be the chief scientist of Songliao Scientific Drilling, the deepest drilling project of ICDP, and now is leading the first IUGS recognized Big Science Program ‘Deep-time Digital Earth’ (DDE).



Roland Oberhänsli was President of the International Union of Geological Sciences IUGS. He graduated in geology in 1973 and earned his doctor degree in Natural Sciences in 1977 at ETH-Zürich in Switzerland. He obtained his *venia docendi* in Petrology and Mineralogy in 1986 at University of Bern.



Michael H. Stephenson is Executive Chief Scientist (Decarbonisation) at the British Geological Survey. He has done research in the Middle East and Asia, including in Oman, Saudi Arabia, Jordan, Pakistan, Iran, Israel and Iraq. He has professorships at Nottingham and Leicester universities in the UK and is a visiting professor at the University of Nanjing, China. He has published three books and over 100 peer-reviewed papers. His recent book ‘Energy and Climate Change: An Introduction to Geological Controls, Interventions and Mitigations’ examines the Earth system science of the formation and use of fossil fuel resources, and the implications for climate change.



Junxuan Fan born in 1973 in Lantian, Shaanxi Province; Ph.D.; graduated from Nanjing Institute of Geology and Palaeontology, Chinese Academy of Sciences; professor at the Nanjing University. He is interested in big data in geoscience, evolution of life, numerical stratigraphy and palaeogeographic reconstruction.