BY YUQI BAI, CLIFFORD A. JACOBS, MEI-PO KWAN,
AND CHRISTOPH WALDMANN

Geoscience and the Technological Revolution

Is geoscience ready to address the challenges and capture the opportunities that will come to the fore over the next decade? Maybe. The imperative for geoscience is to help society understand the Earth system and thus inform decision-making processes. This necessity has never been greater than it is today, nor have the challenges been more complex.

Although the Earth system continuously generates information, our ability to capture this information has increased markedly. Around 2000, it became evident that the quantity, quality, and detail of time and space information were on a path of exponential growth. This is due, in large part, to more complex simulations and more extensive observation systems, all employing advanced technologies. For example, the *Sentinel* satellite system, which is part of the European Commission's Copernicus program, consists of a network of spacecraft and data processing centers that will comprise the largest environmental satellite system in history when fully built. Downlinks from the *Sentinel* satellite system to ground links will transfer information at rates of up to 1.8 Gb/s [1]. High-spatial-resolution climate models can generate petabytes of data from a single set of experiments, with the expectation that exabytes (10^{18} B) of data are possible as models become more sophisticated within the next decade [2].

EMERGING AND DYNAMIC STORIES ABOUT THE FUNCTIONS OF AND INTERACTIONS AMONG EARTH SYSTEMS ARE NOW REGULARLY BEING REVEALED.

Emerging and dynamic stories about the functions of and interactions among Earth systems are now regularly being revealed. Oceanic research, for example, is illuminating the complex changes taking place in the ocean ecosystem caused by the increase in atmospheric carbon dioxide [3]. These stories often cross the disciplinary precincts of geoscience and are increasingly recognized as a coherently compelling tale. This article discusses four specific realities (the deluge of data, coherent access to knowledge, data interoperability, and research reproducibility) for which geoscience is well positioned to avail itself of and exploit the new opportunities the technological revolution creates. At the same time, however, geoscience will face significant challenges in taking advantage of these advances.

DATA DELUGE

Terabyte to petabyte data sets are becoming common, as well as useful, in geoscience. The Earth system is being measured by an ever-growing number of methods including remote and in situ sensors and sophisticated sampling techniques at various locations around the world. In northern latitudes, the European Incoherent Scatter Scientific Association (EISCAT) embarked on EISCAT_3D (Figure 1). The enterprise will create and maintain a multistatic phased-array radar system dedicated to observing the Earth's polar atmosphere and studying how the atmosphere is coupled to space. This project will generate substantial amounts of data: 5 PB per year in 2018 and 40 PB per year in 2023.

Geoscientists astutely use methods, often developed for other purposes, to store, curate, and distribute the vast amounts of collected data and condition these data according to their needs and purposes. Generally, the scientist's intent is to ensure that data and information are openly available, highly accessible, and reliable. But big data is fraught with challenges and missed opportunities [4]. For example, large data sets cannot be shared easily, can be difficult to search, and eschew common formats and structures. This creates impediments to the use and repurposing of data sets. Some of these

Digital Object Identifier 10.1109/MGRS.2016.2635018
Date of publication: 14 September 2017

difficulties are being addressed at data centers through the creation of multinode collaborative environments that support on-site data use by thin clients (i.e., those that do not have a powerful enough computer to connect to and use the information available at the data center; the computing is performed at the center, and the results are displayed on the user's computer). The availability of geographic information systems technologies at data centers is an example of how on-site data use can overcome some of the barriers associated with accessing and using big data [5].

Another challenge, often overlooked, is the long-term archiving and quality assessment required for curation integrity. Earth observation initiatives with a global scope, e.g., the Global Earth Observation System of Systems (GEOSS), identified this issue early on and have proposed solutions such as introducing the concept of community-derived labels that describe the data quality in addition to the traditional attributes provided in metadata; for a broad overview of this topic, see [6].

However, the amount and diversity of data are growing rapidly, and solutions offered today might not be capable of overcoming the underlying challenges associated with big data in the future. Massive—and potentially useful—environmental data sets are being collected through various means, including cell phones and automobiles. However, issues such as geoprivacy violations need to be addressed before access to and use of these data are fully realized, because global positioning system data can disclose the location of people's activities—and even their homes and identities.

FIGURE 1. A view of the EISCAT site. (Photo courtesy of Dr. Tom Grydeland.)

FIGURE 2. Observed and computed information about the Earth system is abundant, but not necessarily available in a coherent and accessible form. (Image courtesy of Yuqi Bai.)

COHERENT ACCESS TO KNOWLEDGE

Global access to geoscience data and knowledge, which has surged as part of the scientific scene over the past couple of decades, offers almost overwhelming opportunities. The availability of this wealth of information is propelled by the establishment of the Internet and web services (Figure 2). Geoscience research is enabled, and hindered, by multifaceted advances in technology. The cultural precincts embedded in the discipline can create barriers to understanding the Earth as an integrated system. Technologies provide new levels of productivity and capability to increase the availability and applications of data. In fact, these have become ingrained in our consciousness as transformative developments for geoscience. At the same time, however, they pose challenges that may limit their promise.

Geoscience lacks systematic practices to expose, access, and classify information and data consistently across all disciplinary areas. Some important data cannot be easily found through commonly used techniques, and other data are restricted for various reasons [7]. (For current efforts on discovering obscure data, see http://deepdiver.stanford.edu/.) A complicating factor is that not all domains in geoscience embrace the same levels of adopted technology, which directly and indirectly affects data availability. Exemplary of these challenges are data collections that reside with small research groups at universities. These data sets are numerous and often highly relevant in their research domain. However, these valuable data sets are disparate and undistributed and lead to a disparity among researchers in terms of information access.

INTEROPERABILITY

Geoscience researchers have no global master plan for developing and adaptating technologies to be used as a means of advancing their studies; nor are there agreed-upon methodologies and practices to be employed in using these technologies. Instead, the choices made, the technologies embraced, and the practices adopted are influenced by the discipline of study, organizational traditions, lack of access to a common cyberinfrastructure, variations in terms of resource availability (human and financial), and other cultural factors, such as accepted practices and user requirements. This has led to a divergent, rather than convergent, set of systems and practices that do not consistently enable efficient and effective interoperability. The driving forces for monsoon events, for example, act on time scales ranging from months to hundreds of thousands of years, and the processes connected to these phenomena make it crucial to take a multidisciplinary approach when studying them, but often these data are frustrating and incompatible (Figure 3).

AN INCREASING AMOUNT OF ENVIRONMENTAL RESEARCH IS NOT REPRODUCIBLE, IF JUDGED BY THE STANDARD OF BEING ABLE TO EXACTLY DUPLICATE AN EXPERIMENT AND ARRIVE AT THE SAME RESULTS.

There are efforts underway to overcome the interoperability dilemma, such as plans for the GEOSS and EarthCube, the latter endeavor sponsored by the National Science Foundation [8]. Progress is limited but positive with respect to understanding and addressing these challenges. However, the most difficult issue to consider is the disparity between *cultural* and *technological* change. The pace of change within the social and cultural context is glacial compared to the current pace of technical innovation. These differences manifest themselves in various ways, such as placing greater value and effort on gathering more data (for which technology is a significant enabler) than on sharing data. The unrelenting advance of widely available technologies adds a factor of complexity. Because more technical choices continually become available, interoperability among various approaches becomes more difficult to maintain. This challenge must be recognized and overcome to fully open up the epic tale of the Earth system.

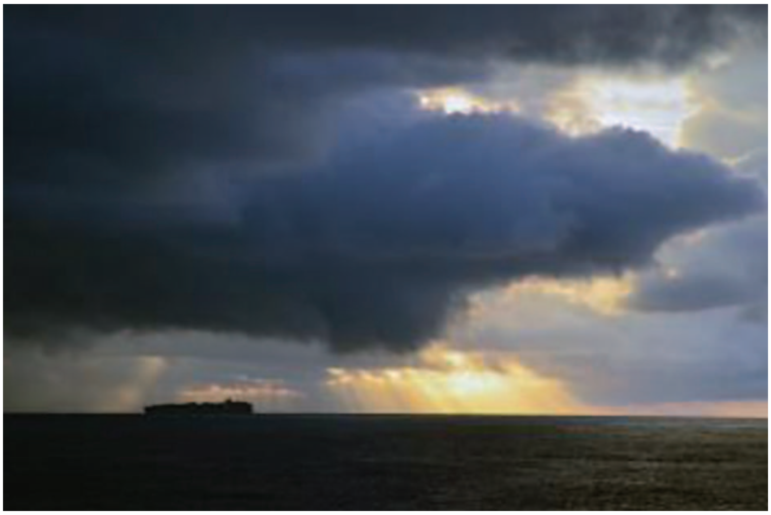

FIGURE 3. A freighter under monsoon rain clouds. (Photo courtesy of M. Mohtadi, Center for Marine Environmental Sciences, Universität Bremen.)

REPRODUCIBLE RESEARCH

An increasing amount of environmental research is not reproducible, if judged by the standard of being able to exactly duplicate an experiment and arrive at the same results. This is because the environment being observed is dynamically changing. Compounding the challenge are the increasingly sophisticated methods used in experiments, which are often not adequately applied and/or documented. Geoscience cannot blithely neglect the spirit of the golden rule of reproducibility, even if the letter of the rule is impossible to enforce under many circumstances. The need to adhere to best principles in the conduct of science is more urgent now than ever.

But there are circumstances under which exact reproducibility of scientific experiments is not possible, such as the bit-wise reproduction of a 100-year Earth system simulation [9] or Eulerian measurements of the upper ocean. It is common for the provenance of the data used in research activity not to be completely known. The fact that scientific experiments cannot be exactly reproduced does not diminish their value, but it does necessitate a qualitative and quantitative assessment of the data. The geoscience community must rethink the appropriate context and standards that should be applied when reproducing the outcomes from research endeavors.

There has never been a greater opportunity to capture and rapidly advance our understanding of the Earth system. This opportunity is created, in part, by the community's ability to harness the ongoing advances in technology and enable their application to achieve societal benefits. There is no doubt that geoscience has harvested the fruits of the technology and knowledge evolution. The discipline believes it is ready and able to tackle future challenges. However, the barriers identified—which constitute only a partial list—become onerous in the milieu of exponential technology and data growth.

For instance, it is relatively easy to creatively apply technology that enables a varied set of observations to be taken simultaneously. The difficulty is in deriving a holistic understanding of the environmental system being observed and identifying the underlying interdependences when there is insufficient knowledge of the processes in play. There is a need for experiment design that ensures its outcomes will be of mutual benefit to all pertinent subdisciplines of geoscience, not just to the group initiating the observations. A more robust discussion about observational experiment design across the geosciences will improve the quality of observations and advance our knowledge of the Earth system.

What we often forget is that science is performed by human beings; thus, it is inextricably linked to a social context. Overcoming the cultural and social impediments to geoscience will further advances in the field, particularly with respect to understanding the dynamic Earth system as an integrated whole. This might be the most difficult obstacle to overcome of all those highlighted, which is why we cannot say for certain that geoscience will be able to address future challenges and opportunities. Perhaps the cultural tipping point is nearly upon us.

There are clear benefits to a social transformation in the conduct of science; it will lead to more productive and capable individuals and institutions. This synergistic integration of geoscience is necessary to provide counsel for the greater good of society. Geoscience must earn the future, because the Earth system changes and society's responses cannot wait.

AUTHOR INFORMATION

Yuqi Bai (yuqibai@mail.tsinghua.edu.cn) received his M.S degree in geographic information systems (GIS) in 2000 from the China University of Geosciences, Beijing, and his Ph.D. degree in cartography and GIS from the Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing, in 2003. His research interests include the use of technologies for geospatial big data management and analytics, geospatial service discovery and integration, high-performance geospatial computing, and standards and policies for large-scale geospatial data sharing and information integration. He is with the Ministry of Education Key Laboratory for Earth System Modeling, Center for Earth System Science, Tsinghua University and the China/Joint Center for Global Change Studies in Beijing, China.

Clifford A. Jacobs (caj2014llc@gmail.com) received his B.S. degree in 1965 in mathematics and his M.S. degree in oceanography, both from Texas A&M University. He received his Ph.D. degree in oceanography in 1976 from New York University. He was at the National Science Foundation for 30 years, with oversight responsibilities in geoscience and computer sciences. He has research experience in modeling and data analysis. He is retired from the National Science Foundation, Arlington, Virginia.

Mei-Po Kwan (mkwan@illinois.edu) received her B.SoSci. degree in 1985 in geography from the Chinese University of Hong Kong, her M.S. degree in urban planning in 1989 from the University of California, Los Angeles, and her Ph.D. degree in geography in 1994 from the University of California, Santa Barbara. Her research interests include environmental health, sustainable cities, human mobility, and the application of geospatial technologies and geoscience and computer science methods in geographic and health research. She is with the Department of Geography and Geographic Information Science at the University of Illinois at Urbana-Champaign and the Department of Human Geography and Spatial Planning at Utrecht University, The Netherlands.

Christoph Waldmann (waldmann@marum.de) received his intermediate diploma in physics from the University of Munich in 1977, his diploma degree in physics from the University of Kiel in 1981, and his Ph.D. degree in physics in 1985 from the University of Kiel. His research interests include the development of new methods of efficient, long-term observation of processes in the sea using a range of techniques from hydroacoustic methods and instruments to optical methods for measuring basic parameters in the water column from mobile and stationary observation platforms. He is with the Center for Marine Environmental Sciences, University of Bremen, Germany.

REFERENCES

[1] S. Clark. (2016, Apr. 28). Europe's Sentinel satellites generating huge "big data" archive. *Spaceflight Now*. [Online]. Available: https://spaceflightnow.com/2016/04/28/europes-sentinel-satellites-generating-huge-big-data-archive/

[2] J. Kinter, T. O'Brien, S. Klein, S. J. Lin, B. Medeiros, S. Penny, W. Putman, K. Raeder, A. Mariotti, and R. Joseph, "High-resolution coupling and initialization to improve predictability and predictions in climate models workshop." U.S. Dept. Energy, DOE/SC-0183; U.S. Dept. Commerce NOAA Tech. Rep. OAR CPO-5, 2016.

[3] K. Bradshaw. (2007, Jan./Feb.). Discovering the effects of CO_2 levels on marine life and global climate sound waves. *U.S. Geological Survey*. [Online]. Available: http://soundwaves.usgs.gov/2007/01/

[4] J. Fan, F. Han, and H. Liu, "Challenges of big data analysis," *Natl Sci Rev.*, vol. 1, no. 2, pp. 293–314, Feb. 2014.

[5] O. Wilhelmi, J. Boehnert, and K. Sampson. (2016, Jan.). Visualizing the climate's future. *Earth and Space Science News*. [Online]. Available: https://eos.org/project-updates/visualizing-the-climates-future

[6] Wikipedia. (2017, June 12). Global Earth observation system of systems. [Online]. Available: https://en.wikipedia.org/wiki/Global_Earth_Observation_System_of_Systems

[7] Z. Ce, "DeepDive: A data management system for automatic knowledge base construction," Ph.D. dissertation. Univ. Wisconsin–Madison, 2015.

[8] National Science Foundation. EarthCube. [Online]. Available: https://www.earthcube.org/

[9] L. Liu, S. Peng, C. Zhang, R. Li, B. Wang, C. Sun, Q. Liu, L. Dong, L. Li, Y. Shi, Y. He, W. Zhao, and G. Yang, "Importance of bitwise identical reproducibility in earth system modeling and status report," *Geosci. Model Dev. Discuss*, vol. 8, pp. 4375–4400, 2015.

GRS