

Introduction to the Special Issue on Geoscience Data Provenance

DATA provenance means the history of a data product, including the sources and processing steps for deriving the product and the ownership of the product. In geoscience, data provenance is also called data lineage. Today, there is an increasing demand to capture provenance in the whole life cycle of remotely sensed geoscience data from acquisition, archival, processing, and distribution, to applications, and to make the provenance information available to the user community. The complete and accurate provenance information can bring transparency to data sharing and production, give credit to data and algorithm contributors, make scientific results reproducible and trustworthy, and support advanced data quality analysis. Motivation for capturing and sharing provenance also comes from the distributed data and information infrastructure that has been benefiting the Earth science community in the last decade, such as Spatial Data and Information Infrastructure, e-Science, and Cyberinfrastructure. In such a data-rich environment, the provenance information is even more important, since scientists rely on the provenance to understand and determine the reliability and usability of a scientific product generated from distributed services and inputs provided by various providers.

In the past decade, geoinformation scientists have conducted a large number of studies on geoscience data provenance. Sponsored by the Data Archiving and Distribution Technical Committee (now known as the Earth Science Informatics Technical Committee) of the IEEE Geoscience and Remote Sensing Society, this special issue, which contains 12 provenance-related papers, represents the state of the art in geoscience data provenance research and applications.

The paper authored by Di, Yue, Ramapriyan, and King, the four guest editors of this special issue, provides an overview of the geospatial data provenance and discusses key considerations and further research issues in the field. There are many models for representing the geoscience data provenance information, notably ISO 19115 and the Open Provenance Model. Feng's paper in this special issue discusses the mapping between the two provenance information models. How to automatically capture the provenance information in data production systems, particularly in the Web service environment, has been the research topic for many provenance studies. The papers authored by Di *et al.* and Jesen *et al.* summarize their work in this area. In Earth Science agencies, there are many operational legacy data systems which are not provenance

aware. The paper authored by Conover *et al.* provides an example of how to introduce provenance capture into legacy data systems. After the provenance information is captured, managing the information is another important area of provenance research. The papers authored by Yuan *et al.*, Huq *et al.*, and Sun *et al.* represent the state of the art in this area. The provenance information is very important in many geoscience applications. The paper authored by Celino discusses how to use the provenance information for the aggregation of volunteered geographic information. The paper authored by Espinoza-Molina and Datcu discusses the retrieval of Earth-observation images by using provenance metadata and other information. The paper authored by Tilmes *et al.* describes how the provenance information is captured, represented, and used in the National Climate Assessment of the United States. The last paper in this special issue, which is authored by Hueni *et al.*, discusses the APEX calibration information system for the generic storage of imaging spectrometer calibration frame data and associated metadata, including the provenance information.

As the guest editors of this special issue, we thank the authors for contributing their papers and the reviewers for their valuable reviews and comments. We hope TGRS readers enjoy the papers in this special issue.

LIPING DI, *Guest Editor*

Center for Spatial Information Science and Systems
George Mason University
Fairfax, VA 22030 USA

PENG YUE, *Guest Editor*

State Key Laboratory of Information Engineering in
Surveying, Mapping and Remote Sensing
Wuhan University
Wuhan 430079, China

HAMPAPURAM K. RAMAPRIYAN, *Guest Editor*

NASA Goddard Space Flight Center
Greenbelt, MD 20771 USA

ROGER L. KING, *Guest Editor*

Center for Advanced Vehicular Systems
Mississippi State University
Starkville, MS 39759 USA



Liping Di (M'01–SM'06) received the Ph.D. degree in remote sensing/geographic information system (geography) from the University of Nebraska, Lincoln, USA, in 1991.

He is with the George Mason University, Fairfax, VA, USA, where he is a Professor and the Founding Director of the Center for Spatial Information Science and Systems and a Professor of the Department of Geography and Geoinformation Science. He has engaged in geoinformatics and remote sensing research for more than 25 years and has published over 300 publications. He has served as the Principal Investigator (PI) for more than \$34 million in research grants and as a Co-PI for more than \$8 million in research grants/contracts awarded by U.S. federal agencies and international organizations. His current research activities are mainly in the following three areas: remote sensing standards, Web-based geospatial information and knowledge systems, and remote sensing applications.

Dr. Di has actively participated in the activities of a number of professional societies and international organizations, such as IEEE Geoscience and Remote Sensing Society (GRSS), International Society for Photogrammetry and Remote Sensing (ISPRS), Committee on Earth Observation Satellites (CEOS), International Organization for Standardization (ISO) Technical Committee (TC) 211, Open Geospatial Consortium, InterNational Committee for Information Technology Standards (INCITS), and Group on Earth Observations (GEO). He served as the Co-chair of the Data Archiving and Distribution Technical Committee of IEEE GRSS from 2002 to 2005, where he became the Chair from 2005 to 2009. He currently chairs INCITS/L1, the U.S. national committee responsible for setting U.S. national standards on geographic information and representing the U.S. at ISO TC 211.



Peng Yue (M'06–SM'12) received the B.S. degree in geodesy and surveying engineering from Wuhan Technical University of Surveying and Mapping, Wuhan, China, in 2000, the M.S. degree in geodesy and survey engineering from the State Key Laboratory of Information Engineering in Surveying Mapping and Remote Sensing (LIESMARS), Wuhan University, Wuhan, in 2003, and the Ph.D. degree in geographic information system (GIS) from LIESMARS in 2007. His doctoral research was conducted at George Mason University, Fairfax, VA, USA, from 2004 to 2007.

He was a Research Associate Professor with George Mason University from 2011 to 2013. He is a Professor with LIESMARS, Wuhan University. His research interests are Earth science data and information system interoperability, Web GIS, geospatial Web Services, and Geospatial Semantic Web. His research has been funded by the National Aeronautics and Space Administration, the National Geospatial-Intelligence Agency, the Department of Energy, Open Geospatial Consortium, the National Natural Science Foundation of China, and the Ministry of Science and

Technology of China. He is the author or coauthor of over 60 publications in refereed scientific journals, conference proceedings, and books.

Dr. Yue is the Secretary of Surveying and Mapping discipline on the Academic Degrees Board of the National Council of China, Secretary of the International Society for Photogrammetry and Remote Sensing Commission VI, and Co-Chair of the IEEE Geoscience and Remote Sensing Society Earth Science Informatics Technical Committee (formerly DAD TC).



Hampapuram K. Ramapriyan (M'03–SM'03) received the B.Sc. degree from the University of Mysore, Mysore, India, the B.E. and M.E. degrees in electrical engineering from the Indian Institute of Science, Bangalore, India, and the Ph.D. degree in electrical engineering from the University of Minnesota, Minneapolis, USA.

He is the Assistant Project Manager of the Earth Science Data and Information System (ESDIS) Project at the Goddard Space Flight Center, National Aeronautics and Space Administration (NASA), Greenbelt, MD, USA. He has over 40 years of managerial and technical experience in science data system development, image processing, remote sensing, parallel processing, algorithm development, science data processing, archiving, and distribution. The ESDIS Project develops and operates one of the largest civilian science data systems in the world—the Earth Observing System Data and Information System (EOSDIS) in support of NASA's Science Mission Directorate. The ESDIS Project has been instrumental in establishing, as a part of EOSDIS, a set of Distributed Active Archive Centers (DAACs) around the U.S. that

manages NASA's Earth science data and provides convenient access to trillions of bytes of data in various scientific disciplines such as land processes, oceanography, hydrology, atmospheric sciences, cryospheric studies, etc. The project has also developed systems that facilitate “one-stop shopping” access to international data centers. His responsibilities in the project have ranged from supervising a group of technical professionals in the design and implementation of EOSDIS and managing the early development and operation of the DAACs to providing a customer focus by interfacing with the scientific customer community to understand their requirements and assuring that the system development accommodates their requirements. He has been involved in the study of the evolution of EOSDIS for the future decade and its implementation. His most recent focus is on data preservation and stewardship. He has developed NASA's Earth Science Data Preservation Content Specification.

Dr. Ramapriyan was the Vice-Chair of the Technical Committee of the IEEE Geoscience and Remote Sensing Society on Data Archiving and Distribution between 2005 and 2009 and its Chair during 2009–2013. He is an active member of Data Stewardship Committee within the U.S. Earth Science Information Partners' Federation.



Roger L. King (M'73–SM'95) received the B.S. degree from West Virginia University, Morgantown, USA, in 1973, the M.S. degree in electrical engineering from the University of Pittsburgh, Pittsburgh, PA, USA, in 1978, and the Ph.D. degree in engineering from the University of Wales, Cardiff, U.K., in 1988.

He began his career with Westinghouse Electric Corporation but soon moved to the U.S. Bureau of Mines Pittsburgh Mining and Safety Research Center. Upon receiving the Ph.D. degree in 1988, he accepted a position with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, USA, where he holds the position of Giles Distinguished Professor and serves as the Director of the Center for Advanced Vehicular Studies (CAVS), Bagley College of Engineering. He also holds the CAVS Endowed Chair in Engineering.

Dr. King has been the recipient of numerous awards for his research including the Department of Interior's Meritorious Service Medal. Over the last 30 years, he has served in a variety of leadership roles with the IEEE Industry Applications Society, Power and Energy Society, and Geosciences and Remote Sensing. He has served for four years as the Chair of the IEEE Geoscience and Remote Sensing Society (GRSS) Data Archiving and Distribution Technical Committee and served as a member of the IEEE GRSS Administrative Committee (AdCom). He also served as the Cotechnical Chair for IGARSS 2009 in Cape Town, South Africa. He is a member of the European Image Information Mining Coordination Group, Tau Beta Pi, Phi Kappa Phi, Sigma Xi, and Eta Kappa Nu.