

## PROCEEDINGS PAPER

# From Geophysical Data to Geophysical Informatics

Fenglin Peng<sup>1,2,4</sup>, Le Peng<sup>5</sup>, Jian Zhang<sup>3,4</sup>, Guoqiang Xue<sup>1,4</sup>, Maining Ma<sup>3,4</sup> and Yunfei Zhang<sup>4</sup>

<sup>1</sup> Key Laboratory of Earth and Planetary Physics, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100029, China  
pengf@geophys.cn

<sup>2</sup> Beijing National Observatory of Space Environment, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100029, China

<sup>3</sup> College of Earth Science, University of Chinese Academy of Sciences, 100039, Beijing, China

<sup>4</sup> Data and Information Platform Center, Committee of Geophysical Informatics, Chinese Geophysical Society, PO Box 9825, Beijing, China

<sup>5</sup> Department of Physics, City University of New York, New York, USA

Geophysics is based on massive data work including data observation, data gathering/collecting, data management, and data analysis. Over the years, in China and other countries, geophysicists and geophysical institutions have accumulated a huge amount of geophysical data, built up many geophysical data banks and data centers, constructed/established many monitoring and transferring systems and infrastructures of geophysical data, and developed many advanced data analysis methods about data on land, ocean, and space. Based on this work, a new discipline of geophysics, geophysical informatics, has been gradually developing over the past 20 years. The recent advances of geophysical informatics in China are introduced and reviewed in this paper.

**Keywords:** Geophysical data; Information science and techniques; Geophysical informatics

## 1 Introduction: Geophysical Data

Over the centuries humankind has studied the Earth, oceans, and the sky. Data were collected in order to explain the physical phenomena of our planet. Understanding tides and earthquakes was challenging because of the difficulty of organizing the results of observations within scientific frameworks in ancient times. Geophysicist pioneers were aware that the accumulation of accurate data was the basis of the development of their scientific theories. By comparing and organizing the values of gravitational acceleration, which were measured at different places, geophysicists developed the Earth's gravitology. The scientists who were focused on geodesy studied the shape of the Earth and the density of the Earth's interior substances. Their work contributed to the definition of international standards for fundamental length units, a basis of physics.

The discovery of geomagnetism relied on the accumulation of the observation data including the intensity and direction of the geomagnetic field. Analysis of the observation data of seismic waves contributed to the foundation of seismology.

Currently, the systematic and comprehensive study of the Earth and geospace has become the focal point for geophysicists. It includes researching the hydrosphere, atmosphere, ionosphere, magnetosphere, terrestrial space, the solar-terrestrial relationship, and planetary geophysics. It also includes the integration of the branches of geophysics such as geomagnetism, geoelectricity, geothermics, geodynamics, and seismology. Many data bank systems were established on theoretical geophysics, environmental geophysics, and exploration geophysics. The geophysical data we have accumulated so far include the data recorded in logbooks and on modern electronic devices.

Some of the data are quantified values. Also, some of the data are literal descriptions of earthquakes and auroras that were recorded in ancient books (historic documents). In the past, scientists used words and pictures to record the details of nature physical phenomena such as earthquakes, auroras, meteors, comets,

and eclipses. Now, scientists record these events by taking photographs and videos. In summary, with the rapid development of information techniques, more and more new types of geophysical data have become available, and the total amount of the geophysical data has increased explosively. These data have become the basis of geophysical informatics.

## 2 Geophysical Observation

Observation data about geophysical phenomena has been accumulating over the centuries. Even during the beginning period of geophysics in China, scientists from Europe, America, and China were working on geomagnetic surveys and related data works in China and surrounding areas (An, Peng, Liu, & Wang, 2014). Nowadays, advanced geophysical observation platforms have become more and more important for the development of geophysics. Those platforms that have multi-dimensions, large space range, high resolution, longer persistent observation time, and shorter time intervals between each observation are necessary for modern geophysics research. These include a national marine survey, a national aerial survey, and a global survey of the Earth, oceans, and space by ships, airplanes, and satellites.

The system also includes various observatories, automatic observation stations, and seafloor seismometers. In addition, the observation platforms built for specific projects such as mining, resource surveys, and engineering surveys supply data for geophysical research. Those observation stations that have been built or will be built on the moon and the planets are also part of the geophysical observation system. These data all need to be transformed, verified, and analyzed. Effectively processing the data is a huge challenge for the modern geophysical observation system.

The development of electronic information techniques and science has provided many effective tools and methods to the observation system. As more and more tools are used by the geophysicist processing the observation data, geophysical informatics develops rapidly.

## 3 The Geophysical Database, The Virtual Platform and their Applications

With the development of information techniques, geophysicists computerized many geophysical databases and networks (Gao, Lu, Wen, & Gao, 1992). Information systems based on the geophysical observatories network and internet databases gradually increased. These databases covered most fields of geophysics, including: (1) basic Earth and planetary physical parameters, (2) geomagnetic data, including data from geomagnetic observatories, geomagnetic surveys, and paleomagnetism, (3) gravity data, (4) geothermics data, (5) geodynamics data, (6) the structure of deep Earth, including data from magnetotelluric sounding and deep seismic sounding, (7) space physics data, including ionosphere and upper-atmosphere physics data, (8) ancient geophysical information and data (Peng, Shen, Tang, et al., 2007). At the end of the 20<sup>th</sup> century, some real-time data platforms were also established, for example, the network of international real-time geomagnetism observatories, INTERMAGNET (Peng, Wang, Zheng, et al., 2007). Attempts were also made to visualize the geomagnetic field in a geographic information system, for example, Google Earth (Wang, Peng, et al., 2008, 2009).

Based on these databases, data centers, and data-information platforms, geophysicists have been able to complete a good deal of research. For example, we have forecast the peak of solar activity by using sunspot data from the World Data Center and geomagnetic data from the Sheshan Observatory. The results of this research project were presented at the International Conference of Solar-Terrestrial Prediction in 1996 (Peng, Chang, Tschu, & Wang, 1997). Other geophysicists have also studied mathematic methods to predict various geophysical phenomena, such as earthquakes, magnetic storms, etc.

## 4 Methodology for Analyzing Geophysical Data and Geophysical Models

Geophysical research has relied on data analysis methods since its inception. These methods include all kinds of analytical as well as mathematical transformations. For example, weather and seismology research have been using the Fourier transformation for a long time, and more recently scientists have worked with wavelet transformation in geophysical research. Because modern geophysics research extracts and analyzes massive amounts of data, more and more advanced tools and methods for data extraction and analysis are needed. These needs prompted the development of information techniques that include tools and methods for data extraction and analysis. For example, the amount of data storage, extraction, and analysis used in the geophysical survey of petroleum exploration was a strong driving force for the development of wavelet transformation.

The need for data analysis methods in researching geomagnetic field models and the gravity field models of the Earth, moon, and the planetary system prompted the development of spherical harmonic analysis.

As a result, this research has benefited from the development of the data analysis method (Papitashvili & Papitashvili, 1991).

## 5 Geophysical Informatics

As with the development of all branches of physics, the development of geophysics was supported by observation, measurements, and philosophical deduction. Geophysical research relies on practical experiments and mathematical analysis. However, unlike most of the other branches of physics, geophysics and planetary physics can hardly collect enough data to derive analytical solutions for their research although the amount of data collected by geophysicists is huge. The objects of geophysical research are complex, and most of them, such as the crust, mantle, and Earth's core, cannot be observed effectively. Therefore, compared with other disciplines in physics, geophysical study needs more information methods and techniques to deal with its data and problems.

During the last two centuries, especially the most recent 50 years, the information techniques and data science in geophysical research have gradually formed a specific academic system. The traditional concept of information techniques and data science in geophysics cannot cover all the contents of this research.

Since the 1980s, geophysicists have begun to discuss a new concept, *Geophysical Informatics*, in their papers. In 1988, a workshop was held by the World Data Center system in Moscow. Geophysical informatics was the main topic of the conference. However, at this conference the content of most of the lectures discussed database techniques only. At that time, the definition of geophysical informatics was not clear (Nechitailenko, 1991).

This new concept was not well known by scientists or even geophysicists until the beginning of 21st century. Since the coming of the new century, information techniques and data science have developed much more rapidly than before. Thus geophysicists are paying much more attention to their applications. In the activities of the 50th anniversary of the International Geophysics Year, held by ICSU (The International Council of Science Unions) and IUGG (The International Union of Geodesy and Geophysics), several of the lectures and academic discussions were about information techniques and data science (Peng, Wu, Gao, et al., 2008; Peng, Zhang, Huang, et al., 2008; Yang, Peng, & Peng, 2008). Since the electronic Geophysical Year, chief officers, senior scholars, and research fellows from the Chinese Geophysical Society, the Chinese committee of the electronic Geophysical Year, the Committee of Geophysical Informatics of the Chinese Geophysical Society, universities, and information science institutes have studied and discussed the definition and the future of geophysical informatics by holding conferences and through other communication channels. An agreement was reached in the matter of how to develop geophysical informatics.

Based on recognition of the contribution of information techniques and data science to the latest developments in geophysics, the authors proposed founding a new discipline, geophysical informatics. During the eGY (electrical Geophysical Year), from July 2007 – Dec. 2008, the main members of the Chinese eGY committee and the international eGY committee supported recognition of this discipline. In the past few years, we have discussed geophysical informatics with many scholars in different branches of geophysics who are in agreement with this concept (Peng, Ma, Peng, Zhang, et al., 2013).

Originally some geophysicists defined the concept of “geophysical informatics” as a volume of data and information about any events, entities, or processes that have existed or continue to exist in the Earth's spherical system. From a planetary geophysics point of view, an Earth's sphere, such as the magnetosphere, ionosphere, atmosphere, etc., could be called an “application domain”, which is limited by the near-Earth space environment. At any moment in time, a number of processes that characterize interactions with planetary Earth's environment are occurring in the sphere. The processes appear in a sphere as an “image”. In this way the “information flow” generated by one sphere characterizes the dynamics and conditions of the physical system.

The databank of the geophysics observation and information system contains data that reflect the physical world. These data can be interpreted as an information system or an information model of the Earth's spheres. The process of mirroring the physical system into an information system is a relationship between the two realities of the world, the generator and the receiver of information. It could be called an information process.

The main attributes of information are structure, meaning, and value. The structure of geophysical information closely connects to the model of the information generator, i.e., the models of the different spheres of the Earth. The structuring of information is conducted in accordance with the entity rules, which connect the data attributes. The value of geophysical information is defined by its substance, completeness, reliability, responsiveness, and a number of other characteristics.

According to the divisions of geophysics, geophysical informatics can be divided into several disciplines, e.g., ionospheric informatics, geomagnetic informatics, etc. Methodologically, for example, in geomagnetic informatics, a geomagnetist could conduct conceptual modeling of geomagnetic data, making a number of requests to the geomagnetism information system and estimating the kinds of movements in the information process (IP) for the perception of the interaction between the magnetosphere and its environment, including the ionosphere. From information science, we should also analyze possible schemes of conceptual elaboration of the knowledge base, which can describe our knowledge about the real world through formalized entities.

In our preliminary summary, geophysical informatics covers the areas listed below:

- 1) Obtaining geophysical data, including the geophysical observation system, survey techniques, and a data transfer system;
- 2) Management of geophysical data, including the geophysical database system and the global network databank grid;
- 3) Research of geophysical data and establishment and study of geophysical data standards;
- 4) Assimilation of geophysical data;
- 5) Analysis and visualization of geophysical data;
- 6) Research of geophysical mathematical models;
- 7) Geophysical prediction, including using traditional mathematical physical methods and new data science methods such as the big data analysis technique;
- 8) Geophysical data mining, including data mining from related or seemingly unrelated fields and from ancient historical documents;
- 9) Integrated geophysical data interpretation.

The above geophysical data properties and operations are the subjects under investigation by geophysical informatics. This means that geophysical informatics is a science studying the principles and transformations of the physical information about the Earth and other planets. With the development of geophysics and data science, the contents described above will be revised to be more and more reasonable and suitable to this discipline in the future.

## 6 Summary

Geophysical informatics research began gradually but has developed quickly over the recent years. It gives geophysics a new motive and powerful tools. By combining new data sciences such as data visualization, big data, and cloud techniques, geophysical informatics will continue to grow.

## 7 Acknowledgements

This work is financially supported by the National Natural Science Foundation of China (40544017), the Chinese Academy of Sciences (XXH12504-1-08, KZZD-EW-TZ-02-02), and the Institute of Geology and Geophysics, Chinese Academy of Sciences Program: Study on geophysical informatics and related infrastructure development".

Also thanks for the financial support from the Ministry of Sciences and Technology of P. R. China. The key project of knowledge creation Chinese Academy of Sciences also gave a lot of support to the development of the WDC system, which was also supported by the Chinese Science and Technical Society.

Thanks for the help from Ms. YANG Jingfeng, Associate research librarian of China Geophysical Data Center; Ms. DONG Jing, Secretary of Chinese Geophysical Society; Ms. SHI Hong, assistant researcher of Beijing Union University.

## 8 References

- An, Z., Peng, F., Liu, S. & Wang, G. (2014) Inspection and study on the geomagnetic survey, charts and models during 1683–1949 in China. *Chinese Journal of Geophysics* 57(11), pp 3795–3803, doi: <http://dx.doi.org/10.6038/cjg20141133>
- Gao, M. Q., Lu, W., Wen, X., & Gao, M. (1992) China Geophysics Database in WDC-D for Geophysics. *CODATA Bulletin* 24(6), p 2.
- Nechitailenko, N.A. (1991) Distributed Information System for Planetary Geophysics. *Proceedings of the Workshop on Geophysical Informatics*, Moscow, Aug. 14–18, 1988, p 54.
- Papitashvili, N.E. & Papitashvili, V.O. (1991) Dynamic Model of Geospheres in Geophysical Informatics and the Approach to the Knowledge Base on Planetary Geophysics. *Proceedings of the Workshop on Geophysical Informatics*, Moscow, Aug. 14–18, 1988, p 89.
- Peng, F., Chang, H., Tschu, K. K., & Wang, J. L. (1997) The Use of Geomagnetic AQD near Sq Focus for Predicting the Magnitude of Sunspot Maximum. *Solar-Terr. Predict.* 5, pp 99–102.
- Peng, F., Ma, M., Peng, L., Zhang, J., Chen, G., Li, Y., Sun, B., & Zhang, Y. (2013) Activities and Plan of the Center for Geophysics (Beijing) from WDC to WDS. *Data Science Journal* 12, p WDS52.
- Peng, F., Shen, X., Tang, K., Zhang, J., Huang, Q., Yue, B., & Wang, D. (2007) Data-sharing Work of the World Data Center for Geophysics, Beijing. *Data Science Journal* 6, pp 404–407.
- Peng, F., Wang, D., Zheng, X., Xing L., Tang, K., Yue, B., Shen, X., Zhang, J., Peng, L., & Huang, Q. (2007) The Internet Databases of the World Data Center for Geophysics, Beijing. *Data Science Journal* 6, pp 879–882.
- Peng, F., Wu, Z., Guo, J., Zhang, J., Chen, X., Zhu, R., Wang, S., & Liu, G., (2008) eGY in China: From IT to geophysical informatics and its outreach. *21st International CODATA Conference*, Kyiv, Ukraine, p 139.
- Peng, F., Zhang, J., Huang, Q., Chen, X., & Guo, J. (2008) Advance of Informatization of International Earth-Space Research and Geophysical Informatics. *Chinese Geophysics* 1.
- Wang, D., Peng, F., Ma, M., Yuan, X., Bai, C., & Sun, L. (2009) Visualization Research of IGRF Model. *Seismological and Geomagnetic Observation and Research* 30(4), pp 7–11.
- Wang, D., Shen, X., Peng, F., Yuan, X., Tang, K., Xing, L., & Peng, L. (2008) Basic Research of IGRF 10 Model on Web Visualization. *21st International CODATA Conference*, Kyiv, Ukraine, p 51.
- Yang, L., Peng, L., & Peng, F. (2008) The Weather, Geologic and Geophysical Environment as a Factor in Transportation Information System. *21st International CODATA Conference*, Kyiv, Ukraine, pp 157.

**How to cite this article:** Peng, F, Peng, L, Zhang, J, Xue, G, Ma, M and Zhang, F 2015 From Geophysical Data to Geophysical Informatics. *Data Science Journal*, 14: 13, pp.1-5, DOI: <http://dx.doi.org/10.5334/dsj-2015-013>

**Published:** 22 May 2015

**Copyright:** © 2015 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License (CC-BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/3.0/>.

 *Data Science Journal* is a peer-reviewed open access journal published by Ubiquity Press.

**OPEN ACCESS** 