# KEY RESEARCH AND APPLICATION OF THE REAL-TIME SERVICE OF REMOTE SENSING BIG DATA

Xinyan. ZHENG [1], Tao. CHENG [1] [*], Zhenyu. LI [2], Chen. Chen [3], Weiwei ZHANG[4], Jin. LIU [1], Xuchao. Xing [1],Wei. HUANG [1],
Zhuhua. ZHA [1], Hongping. ZHANG [1], Dejin. TANG [1], Cong. WANG [1], Heng. LI [1], Lei. DING [1], Zhen. WANG [1], Xinyuan. GAO [3],

1 National Geomatics Center of China, Beijing, China, xinyan_zheng@ngcc.cn, chengtao@ngcc.cn, liujin@ngcc.cn,
xingxuchao@ngcc.cn, huangwei@ngcc.cn, zhazh@ngcc.cn, zhanghongping@ngcc.cn, tangdejin@ngcc.cn, wangcong@ngcc.cn,
liheng@ngcc.cn, dinglei@ngcc.cn, wangzhen@ngcc.cn
2 ADASpace Science and Technology Co.Ltd, Chengdu, China, lizhenyu@national-space.com
3 Land Satellite Remote Sensing Application Center, MNR, Beijing, China, 410208288@qq.com, gaoxy@lasac.cn
4 School of Geography Science and Geomatics Engineering, Suzhou University of Science and Technology, Suzhou, China,
zww9820@126.com

**Commission III, WG III/6**

**KEY WORDS:** Remote Sensing, Big Data, Real-time service, Multi-source, Multi-scale, Multi-spectral. Spatiotemporal Data Management, Big remotely sensed data.

**ABSTRACT:**

With the rapid development of earth observation technology, the remote sensing data gradually becomes the significant data resources for spatiotemporal data analysis. However, to realize the efficient management and quick real-time service of the multi-source, multi-scale and multi-spectral is the key problem for the data management units. This paper brings up a novel technology system and resolution for online and real-time service of the massive remote sensing images by innovatively developing a data management model based on an extensible discrete global grid, inventing the grid-based search and dispatch method for remote sensing big data, firstly building a parameter-driven dynamic map service mode for multi -source, multi-scale and multi-spectral remote sensing big data. This method has greatly improved the efficiency for data management units, such as the Chinese government, corporations etc. and has produced a lot of social and economic benefits.

## 1. INTRODUCTION

Big data, which involve tremendous social, economic and scientific value, is considered as the oil of the future world. It has become the hot spot in the world of the business, technology and politics. In 2008 and 2011, the special issues of the research on the big data have been published in succession (David, 2008; Wouter et al., 2011), which means the coming of the era of the big data.

In the realm of remote sensing and earth observation, with the development of the earth observation technology, the capability of the mankind's comprehensive observation of the earth has reached an unprecedented level. With abundant spatiotemporal information within the remote sensing data, a variety of applications of the massive remote sensing data have been utilized in multiple scenarios such as agriculture, meteorology and emergency response etc. (Huang et al., 2018; Wang et al., 2013; Mahmoud et al., 2013) The remote sensing data presents obvious characteristics of big data (Li et al., 2014). The amount of remote sensing image data increases significantly and exponentially; the speed of data acquisition is accelerated, the update cycle is shortened, and the timeliness is stronger and stronger; the remote sensing data is increasingly diversified with the coexistence of the data with different imaging methods, different bands and resolutions. However, the remote sensing information processing capacity is very low, compared with the acquisition capability of the remote sensing (Li et al., 2012; Quartulli et al., 2013).

For remote sensing big data, one of the most important US government projects is the Earth Observing System Data and Information System (EOSDIS). It provides end-to-end capabilities for managing NASA's Earth science data from various sources. In Europe, "Big Data from Space" conference was organized by the European Space Agency in 2017 (Liu et al., 2018).

With the abundant application scenarios of remote sensing big data, the effective management and service of the remote sensing data has become the key point in the realm of the application of the remote sensing data (Chi et al., 2016; Lin et al., 2013). The way to effectively storage and manage remote sensing big data is a bottle-neck problem which has become the ceiling of the traditional commercial software. This paper focuses on solving the problem of the storage, management and real-time online service of the multi-source, multi-scale, and multi-spectral remote sensing big data on the basis of the extensible mesh division rule. By this extensible mesh division rule, the logic interval level between the remote sensing data and the application has been built, and the index and search of the information among the remote sensing big data gets more effective and flexible compared to the traditional file-based storage of remote sensing data, and the real-time online service becomes possible with massive remote sensing data.

---

* Tao CHENG: chengtao@ngcc.cn

## 2. THE TRADIONAL COMMON METHOD OF THE REMOTE SENSING BIG DATA PROCESS AND MANAGEMENT

With the rapid development of earth observation technology, the remote sensing data gradually becomes the significant data resources for spatiotemporal data analysis (Wang et al., 2015). Distributed file system, with the advancement of the data storage capability, efficiency and cost, has gradually developed into a mainstream remote sensing storage scheme (Wu et al., 2020).

The process of one image or several images covering the gross images from the satellite to service online could be done by one or several traditional commercial software without taking too much time. However, the volume of huge multi-source, multi-resolution and multi-temporal geospatial data goes to TB or PB covering the whole China territory with 9,600,000 km$^2$ for more than one year. Therefore, the storage, backup, building the pyramids, and releasing these huge geospatial data with amount of TB or PB by traditional merchandized software, even the internationally famous software, is still a huge amount of work, which will usually takes several months annually for the images of one year covering the whole China.

With the rapid development of remote sensing technology, the resolution of the remote sensing images has been continuously improved, and the volume of a single image is constantly multiplying. Ever since 2015, the volume of a single remote sensing image data with the resolution better than 1m is between 3GB to 14 GB, and the overall remote sensing images of the project of China's first national geographic conditions census and monitoring is more than 300 TB each year. Thus, as time goes by, from 2015 to nowadays, the overall volume of remote sensing image data has reached PB with multi-source, multi-resolution and multi-temporal geospatial data covered the whole country each year. One single operation, such as stretching, which is just one of the data processing and management of the remote sensing images by the traditional business software, is to process every single point of the whole image (Fig. 1). Thus, one single stretching process on the remote sensing of the whole China in 2015 will have to open, scan and process every single point of over 20, 000 images, which will take a long time. Thus The whole process of stretching, mosaicking, slicing and releasing etc. of the remote sensing images of the project of China's first national geographic conditions census and monitoring for one year will take at least several months. If one singe parameter changes, the whole images will have to be sliced from the beginning again, and the processing time and storage space will exponential grow. In order to efficiently process and manage remote sensing images, and realize the quick and dynamic service, our team has brought up a real-time technical method based on the mesh division which systematically changed the process and manage method of the massive remote sensing images.

The procedure of managing the huge multi-source, multi-resolution and multi-temporal geospatial data of the traditional merchandised software is: 1) Build pyramids for every image of the whole geospatial data; 2) Process huge geospatial data based on these pyramids, including removing the black edges of the images, stretching the images, building the mosaic datasets, unifier ray and colour of the images etc.; 3) Publish each one of them into WMTS service. Since the process is based on building the pyramids, the more huge the data, the more time will be taken. When the devices are limited within 5 computers,

the time will be taken for process the huge multi-source, multi-resolution and multi-temporal geospatial data will be over six months. The storage, backup, processing and releasing these huge geospatial data with amount of PB by traditional merchandized software, even the internationally famous one, is still a huge amount of work.

Many division models of the earth surface have been proposed till now and among them, the equal latitude/longitude division model is not only the simplest, but also highly efficient in computation and storage, making it widely used in the engineering field (Deng et al., 2003). In the equal latitude/longitude division model, interweave equally on the earth surface, resulting in a grid of fixed size. The multi-level grids formed by the equal latitude/longitude division model are actually a multi-resolution pyramid structure.
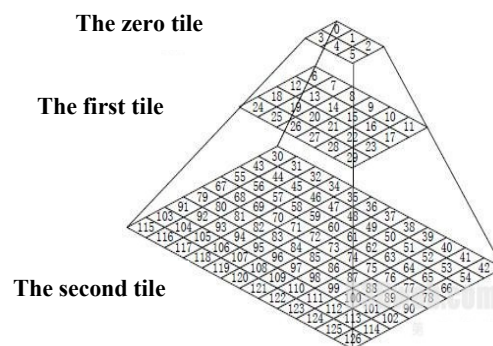


**Figure 1.** The traditional process method of remote sensing images

## 3. THE KEY RESEARCH OF REAL-TIME TECHNICAL METHOD BASED ON EXTENSIBLE DISCRETE GLOBAL GRID

The completely new method for the process and management of massive and big data of the remote sensing images, which is based on the partitioning theory, has developed a series of key technique that has successfully has changed the traditional mosaic-pre-slice image process. A new application mode, "storage is management, data is service, and what you get is what you need", has become a possible and efficient way for image process, management and real-time service.

### 3.1 DATA MANAGEMENT MODEL BASED ON AN EXTENSIBLE DISCRETE GLOBAL GRID

Modern environmental monitoring and modeling requires partitioning the earth's surface into a global grid optimized for survey sampling and unbiased, spatially complete data collection of relevant environmental phenomena (Goodchild et al., 2002). Scientists often favour develop a hierarchical, geometrically regular global partitioning system that is unbiased with respect to spatial patterns created by natural and human processes (White, 1998). The geometrically ideal global partitioning grid would consist of grid cells equal in surface area and identical in shape, akin to the square (Goodchild et al., 2002). The Discrete global grid is an excellent model to simulate the real earth's surface by subdivision infinitely and hierarchically with grids. The grids with the characteristics of multi-resolutions could help to analyse the problems of the same spatial position at different accuracy well (Dutton, 1999; Lukatela, 2000).

According to the remote sensing big data with multi-source, multi-scale, multi-spectral, the key to solve the efficient problem is to design the data management model. As the storage space and resources are limited based on the frugality principle for the Chinese government these days, a novel data management model based on an extensible discrete global grid is brought up, which a grid layer between the data layer and service layer. This grid layer is a three-dimensioned data management framework based on an extensible discrete global grid consists of the basic spatial grid, basic time grid and spectral grid (Fig. 3). The spatial grid in our resolution is an extensible discrete global grid which is compatible with commonly famous grids, such as WGS-84 tile grid, Web Mercator grid, Tianditu grid, GEOSOT grid etc. (Fig.4)
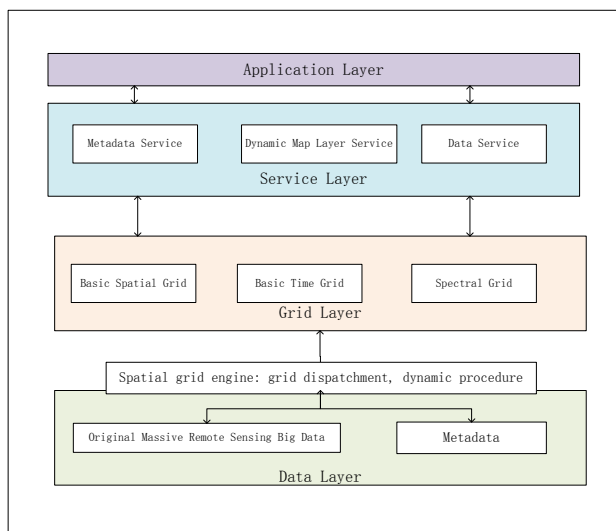


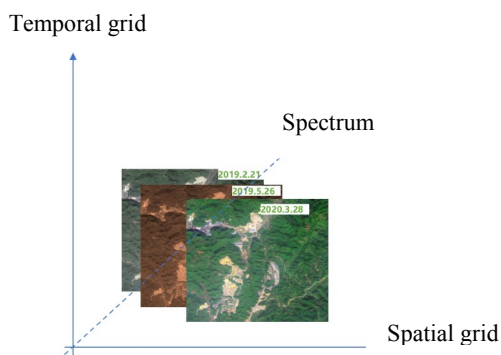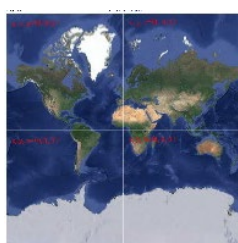**Figure 2. Data organization model based on an extensible discrete global grid**
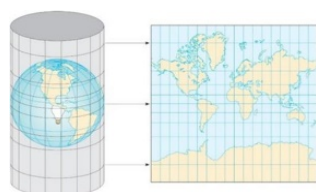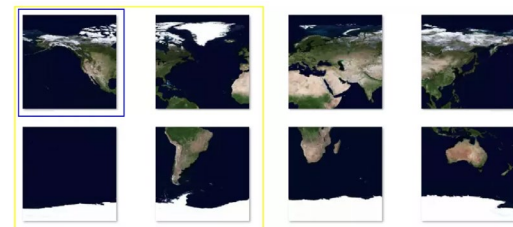


**Figure 3**. The new data management model
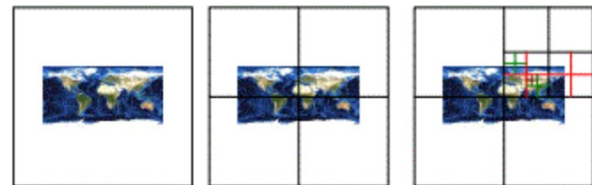


(a) WGS84 tile grid          (b) Web Mercator grid



(c) Tianditu grid



(d) GEOSOT grid

**Figure 4.The extensible discrete global grid**

The basic spatial grid adopts the global partitioning grid which could subdivide hierarchically from global scale to the more detailed scale. A collection of global portioning grids is established and a unique coding for every single grid is assigned by mathematical method. A uniformed global spatial index of remote sensing big data is established which is the basis for data management and spatial search. The basic time grid is an index based on the years, seasons and months. The spectral grid is the index based on the professional remote sensing spectrums such as red, green, blue, infrared, and panchromatic etc.

Traditionally, every single operation of the remote sensing images will trigger the traversing process of the whole file over the 20,000 images of one year, which takes huge amount of time and computing resources. The establishment of a global spatial mesh division index and coding system will help to locate the file and locate the point far more efficiently, compare to the traditional traversing mode of the common business software.

With this model, the original remote sensing data, which is of huge volume, is stored and stay the same all the time during the whole procedure. During the process of remote sensing big data, the intermediate data will just take less than 1% of the storage space which will save a huge amount of storage space compared to the traditional methods.

When searching and displaying the remote sensing big data as in the red rectangle (Fig. 5), the spatial grid engine firstly positions the coordinates of the minimum bounding rectangle (short for MBR), and secondly calculate in the global spatial index to locate the four images that is included in the red rectangle. In order to search and display the part ① in the left upper image, the grid-oriented data dispatch method (Fig. 6) would parse the beginning of the file, and build the logic row-column model. After resolving the physical addresses by grids and read and display just the part ①, the other parts of the left upper image will not be processed and displayed multiple times when the mouse choose the red rectangle. The same parsing,

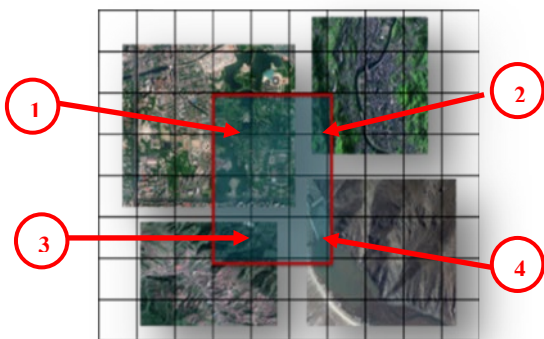reading and displaying way will be utilized in the part ②, ③ and ④.



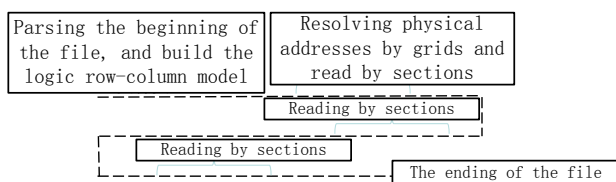**Figure 5. Search and display the remote sensing big data**



**Figure 6. The grid-oriented data dispatch method**

Although, this method will be almost the same when processing just one or two remote sensing images with the data volume of 3GB to 14GB. The more huge the data volume, the more difference will the efficiency will be. When the data volume gets to PB, the efficiency is much quicker compared to the traditional merchant software. Take the remote sensing big data of 500TB for example, the data includes approximately 50,000 images with the spectrum of 4 bands and the resolution better than 1m. It will take at least six months for at least eight persons with five server to accomplish the whole process of removing the black edges, dynamic stretching, dynamic mosaicking, time filtering and releasing the service. Even some of the dynamic processing is not supportive by the traditional merchant software (Tab. 1).

|  | Traditional Merchandized Software | Gid-oriented Dispatch Method |
|---|---|---|
| Dynamic Stretching | Not supportive | supportive |
| Dynamic Mosaic | Not supportive | supportive |
| Time Filtering | Not supportive | supportive |
| Release Time (with 5 server) | At least six months of eight persons | Less than half an hour |
| The Intermediate Data volume | At least 150TB | Less than 100MB |

**Table 1. The Comparison between Traditional Merchandized Software and the Gid-oriented Dispatch Method**

## 3.2 PARAMETER DRIVEN DYNAMIC ONE-MAP OF REAL-TIME SERVICE OF MASSIVE MULTI-TEMPORAL REMOTE SENSING IMAGE DATA

The parameter driven dynamic one-map of real-time publishing service mode of massive multi-temporal remote sensing image data is firstly brought up. This technique (Fig.7) is a unified dispatching technique of multi-temporal remote sensing images based on dynamic spatial-temporal search elaborated in the data management model. All the multi-temporal data is simulated as a virtue layer indexed in the extensible discrete grid, where the

data content could be generated by real time calculation. The spatial calculation is indexed in the grid, and the time sorting could accomplish according to the time information in the metadata. When published as a unified service, a virtual WMTS layer of a front end is dispatched all the time, while the browse request will be responded by the dynamic spatial-temporal search on the server side. A group of map tiles will be dynamically mosaicked according to the time logic, which effectively solves the problem of performance bottle-neck of service invocation of multi- temporal layers. As the platform (Fig.8) could quickly search the remote sensing big data, display all the images that cover the windows area, and put the newest image on the top with a timeline.
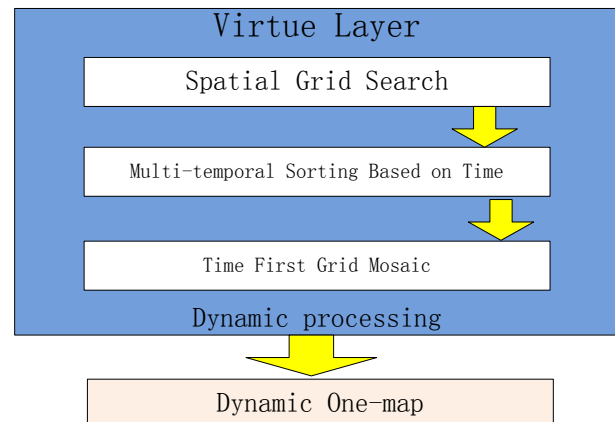


**Figure 7. The solution for dynamic one-map service**



**Figure 8. The platform for real-time service of dynamic one-map**

## 3.3 DYNAMIC IMAGE FUSION TECHNIQUE OF PANCHROMATIC AND MULTI-SPECTRAL IMAGES

The dynamic image fusion technique of panchromatic and multi-spectral images could keep the original data files without creating new data files, and dynamically resample the panchromatic and multi-spectral images by RGB pictures of 3 bands. The traditional method would need the preprocessment for the fusion images which will take a great amount of work and huge storage space. With this technique, the data process time could be saved by 20%, and the data storage space could be reduced by 60%.

**Figure 8. The dynamic fusion of panchromatic and multi-spectral images**

## 4. THE APPLICATIONS OF REAL-TIME SERVICE OF MASSIVE REMOTE SENSING IMAGE DATA

This real-time service has been applied already in several different departments of the central government, such as Ministry of Natural Resources, Ministry of Water Resources, etc. and has played an important role in massive remote sensing data storage, management and visual representation. For example, the technology and platform has successfully helped the Chinese government to make decisions wisely by releasing remote sensing big data from more than thirty different types of satellites with the volume of 200, 000 images as the maps on the China Central Television (Fig. 7). This technology and platform has helped to release 8,000 to 10,000 remote sensing images every day online which means to produce, release and apply all on the same day for the Ministry of Ecology and Environment of the People's Republic of China.

By utilizing this technology, the cumbersome processes for the remote sensing big data with multi-sources, multi-scales and multi-spectral are greatly simplified. This has saved huge space for storing and processing, which also brings huge economic and social benefits.



**Figure 7. The real-time service of massive remote sensing image data**

## ACKNOWLEDGEMENT

The quick and dynamic real-time service of massive remote sensing image data, based on mesh division, has been an important and difficult hotspot for the data management departments for a long time. The advancement of this technique helps to save a lot of storage space and time for the data process and management, which is more and more important for the departments with the annually added data volume of TB or even bigger.

## REFERENCES

CHI M., PLAZA A., BENEDIKTSSON J.A., et al. Big Data for Remote Sensing: Challenges and Opportunities. Proceedings of the IEEE, 2016, 104 (11): 2 207 – 2 219.

DAVID G, 2008．BigData．Nature, 455(7209): 1-136.

DENG X Q. Research of service architecture and algorithms for grid spatial data (in Chinese). Dissertation of Doctoral Degree. Zhengzhou: University of PLA Information Engineering, 2003.

DUTTON G. H. A Hierarchical Coordinate System for Geoprocessing and Cartography. Berlin Heidelberg: Springer-Verlag, 1999.

GONG J Y, XIANG L G, CHEN J, et al. Multi-source geospatial information integration and sharing in Virtual Globes. Sci China Tech Sci, 2010, 53(Suppl Ⅰ): 1-6, doi: 10.1007/s11431-010-3216-0.

GOODCHILD M.F., KIMERLING, A. J. Discrete Global Grids: A Web Book. 2002.

HU Bailin, WANG Chunbo, ZHAO Xuesheng,et at. Degenerate Quadtree Latitude/Longitude Grid Based on WGS-84 Ellipsoidal Facet. Acta Geodaetica et Cartographica Sinica, 2016, 45 (S1): 32-39. DOI: 10.11947/j. AGCS. 2016. F004.

LI Deren,TONG Qingxi,LI Rongxing,etal, 2012．Some Frontier Problems of High Resolution Earth Observation．Scientia Sinica Terrae, 42(6): 805-813.

LI Deren, ZHANG Liangpei, XIA Guisong, 2014．Automatic Analysis and Mining of Remote Sensing Big Data．Acta Geodaetica et Cartographica Sinica, 43(12): 1211-1216. DOI: 10.13485/j.cnki.11-208.2014.0187.

LI Shuang, CHENG Chengqi, TONG Xiaochong,et al, 2016．A Study on Data Storage and Management for Massive Remote Sensing Data Based on Multi-level Grid Model. Acta Geodaetica et Cartographica Sinica., 2016, 45（S1）：106-114. DOI: 10.11947/j.AGCS.2016.F013.

LIN F. C., CHUNG L. K., Ku W. Y., et al. The Framework of Cloud Computing Platform for Massive Remote Sensing Images. IEEE International Conference on Advanced Information Network & Applications, Barcelona, Spain, 2013.

LIU Peng, DI Liping, DU Qian, et al. Remote Sensing Big Data: Theory, Methods and Applications. Remote Sensing. 2018, 10, 711; doi: 10.3390/rs10050711.

LUKATELA H. Ellipsoidal Area Computations of Large Terrestrial Objects. Proceedings of the 1$^{st}$ International

Conference on Discrete Grids' 2000. Santa Barbara: [s. n.] , 2000.

QUARTULLI M, OLAIZOLA I G, 2013. A Review of EO Image Information Mining．ISPRS Journal of Photogrammetry and Remote Sensing, (75): 11-28.

YUAN Fajie. A New Strategy of Storage & Retrieval for Massive Tile Data of Remote Sensing Images. Chengdu: University of Electronic Science and Technology of China, 2013.

WANG C., Hu F., Hu X., et al. A Hadoop-Based Distributed Framework for Efficient Managing and Processing Big Remote Sensing Images. ISPRS Ann, 2015, 2(7): 63 – 66.

WANG P., WANG J., CHEN Y., et al. Rapid Processing of Remote Sensing Images Based on Cloud Computing. Future Generation Computer Systems, 2013, 29(8): 1 963 – 1 968.

WOUTER L, John W, 2011. Dealing with Big Data．Science, 331(6018): 639-806.

WU Huayi, CHENG H., ZHENG J. et al. RS-ODMS: An Online Distributed Management and Service Framework for Remote Sensing Data. Geomatics and Information Service of Wuhan University, 2020, 45(12): 1836-1846. DOI: 10.13203/j.whugis20200198.