

Research on New Spatial Data Infrastructure Supports the Circulation of Geographic Information Data Elements

Heng Li^{1,2,3}, Wei Huang¹, Yong Zhai¹, Wenhao Zhao¹, Xinyan Zheng¹, Jin Liu¹

¹National Geomatics Center of China, Beijing, China-(liheng, huangwei, zhaiyong, zhaowh, xinyan_zheng, liujin)@ngcc.cn

²Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China

³School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China

KEY WORDS: Geospatial Data, Spatial Data Infrastructure, Data Circulation, Data Elements, Digital China.

ABSTRACT:

Data is an important factor of production in the digital economy era and an important support for building a new development pattern. As an important part of data resources, geographic information data is also an important new factor of production, which plays a unique role in the construction of digital China and economic and social development. From the perspective of releasing the value of data elements, data infrastructure is a new type of infrastructure that provides integrated data aggregation, processing, circulation, application, operation, and security services for society under the support of network, computing power and other facilities. It is an organic whole covering hardware, software, open source protocols, standards, and mechanism design. Spatial data infrastructure (SDI) is a collection of technologies, policies and systems that realize the access and access of spatial data. Through the construction of new spatial data infrastructure (NSDI), it supports the circulation of geographic information data elements, making it an important part of data infrastructure. Firstly, this paper gives the connotation of geographic information data elements circulation. Secondly, it proposed to construct the circulation system of geographic information data elements on the basis of enumerating and summarizing the spatial data infrastructure, in order to better support the construction of digital China and economic and social development. Thirdly, the basic requirements, technical indicators and overall architecture of NSDI are designed. Finally, the opportunities and challenges faced by the circulation of geographic information data elements were summarized, mainly focusing on the aspects of policy mechanisms and security systems. At the end of the paper is a summary and acknowledgments.

1. Introduction

Recently, the National Data Administration and 17 other departments jointly issued the three-year Action Plan of "Data Elements ×" (2024-2026) (hereinafter referred to as the "Action Plan"), which takes the high-level application of data elements as the main line, and focuses on promoting the collaborative optimization, reuse and efficiency enhancement of data elements, and the integration and innovation of data elements, and strengthening the scene demand pull. We will promote high-quality supply, compliance and efficient circulation of data elements, and foster new industries, new models, and new drivers of growth. Giving full play to the multiplier effect of data factors will provide strong support for promoting high-quality development and promoting Chinese-style modernization.

Improving data supply is an important guarantee to play the multiplier effect of data factors, and the effective release of public data value is an important source. The amount of public data is large and the quality is good, and its opening is of great significance to the development of the data factor market. Geographic information data is also an important new factor of production. Spatial data infrastructure (SDI) is a collection of technologies, policies and systems that realize the access and access of spatial data. Through the construction of new spatial data infrastructure (NSDI), it supports the circulation of geographic information data elements, making it an important part of data infrastructure.

This paper gives the connotation of geographic information data elements circulation at first. Then, it proposed to construct the

circulation system of geographic information data elements on the basis of enumerating and summarizing the spatial data infrastructure, in order to better support the construction of digital China and economic and social development. Thirdly, the basic requirements, technical indicators and overall architecture of NSDI are designed. Finally, the opportunities and challenges faced by the circulation of geographic information data elements were summarized, mainly focusing on the aspects of policy mechanisms and security systems. At the end of the paper is a summary and acknowledgments.

2. Related Work

The research on the circulation of geographic information data elements by domestic and foreign scholars mainly focuses on two aspects: the open sharing of geographic information data and the construction and application of spatio-temporal big data system platform.

Open Sharing of Geographic Information Data: The United States first signed a memorandum of Understanding on Transparent and Open Government in 2009 and issued an "Open Government Plan". In 2011, the "Geospatial Platform" plan was launched. The UK launched the "Federal Government Data Sharing scheme" in 2010. In addition, Canada, Australia, Japan, South Korea, Singapore and other countries have built a series of open data sharing platform websites. China launched the National Geographic Information Public Service Platform (TIANDITU) in 2010 to promote the open sharing of geographic information data. Li et al. proposes to build an open sharing platform for spatiotemporal data, and gives the basic requirements of four parts: spatiotemporal benchmark,

spatiotemporal modeling big data, cloud computing-based spatiotemporal big data system and supporting environment in 2022.

Construction and Application of Spatio-temporal Big Data System Platform: In the mid-1980s, spatio-temporal databases and their query languages appeared. Since the 21st century, with the proposal of the concept of "spatio-temporal big data", Li Deren, Wang Jiayao and other academicians have used spatio-temporal big data to mine, analyze and visualize. Harvey et al. proposed Geographic Data Mining and Knowledge Discovery, 2015. Academician Li Deren first proposed the realization of global socio-economic dynamic monitoring supported by night-light remote sensing marketing based on remote sensing big data. At present, the use of spatio-temporal big data system platform to build digital city and smart city to improve the level of city management has been widely constructed and applied in the world. Li et al. proposes to the construction of geographic information data circulation system, designed the overall framework, basic requirements and technical indicators of geographic information data infrastructure and geographic information resource database, in order to better support the construction of Digital China in 2023.

3. The Connotation and Applications of Geographic Information Data Elements Circulation

The circulation of geographic information data elements can be in many forms, such as sharing exchange, development and utilization, and trading circulation.

3.1 The Data Sharing Exchange

Take the data sharing exchange as an example. Article 24 of the China Data Security Law clearly points out that the state shall formulate the government affairs data open catalogue, build a unified and standardized, interconnected, safe and controllable government affairs data open platform, and promote the open utilization of government affairs data, the technical architecture as shown in Figure 1.

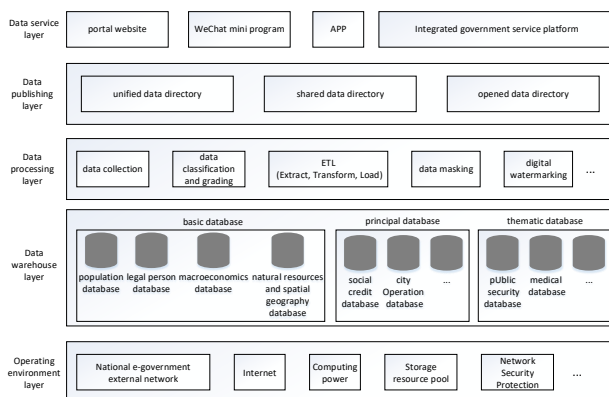


Figure 1. The technical architecture of data sharing exchange.

Infrastructure Layer: Relying on e-government external network, Internet and other network environments, it has certain computing power and storage resource pool, as well as corresponding security protection capabilities.

Data Warehouse Layer: It collects four basic databases including population, legal person, macro economy, natural

resources and spatial geography, theme databases such as social credit and urban operation, and thematic databases such as public security, medical treatment and social security.

Data Processing Layer: For the centralized storage of database information resources, data collection, classification and grading, ETL (including extraction, transformation, loading) operation, data desensitization, digital watermarking and other data processing, to form a data source for open sharing.

Data Publishing Layer: Form a unified data directory based on the data source, which can be shared and opened.

Data Service Layer: Providing data services to government departments, enterprises, institutions and the public by relying on government portal websites, wechat applets, apps and integrated government information sharing and exchange platforms.

3.2 The Data Development and Utilization

Raw data is really not open to share, it needs to be made into data products, there are development costs, so commercial development and utilization is conducive to mobilize enthusiasm. In addition, some public data are not needed by the majority of citizens and market players. We use public costs and taxes to develop these data for a few enterprises to use, and users pay. This kind of public service is a reasonable option, because it only serves the commercial use of a few people.

Pull data through application program interface, this mode allows different applications or systems to exchange and integrate data to achieve data flow and sharing. Generally, the API interface is used (API mode for short). API mode is widely used in data circulation, it can help realize data sharing, data exchange, data migration, data synchronization and other functions.

3.3 The Data Trade

Data trading does not avoid "ownership", but more emphasis on the right to hold, use, and management rights. "Data 20" proposes to explore the structural separation system of data property rights, so as to adopt the structural separation system in the overall framework, and adopt the classification and authorization right of authorization in the specific operation. It creatively proposes the establishment of the data property rights system framework of the "three rights separation" of data resource holding right, data processing and use right and data product management right, and constructs the data property rights system with Chinese characteristics. Data can be confirmed and evaluated into the table to become data resources and data assets. "Data assets" can increase corporate credibility and improve balance sheets, and can play a financing role.

Recently, the Department of Natural Resources of Guizhou Province and the Bureau of Big Data Development of Guizhou Province jointly issued the Pilot Work Plan for the Marketization of Spatial and Temporal Big Data of Geographic Information in Guizhou Province. The "Plan" is the first work plan of the country on the circulation and transaction of spatial and temporal big data of geographic information. The issuance of the "Plan" provides important support for releasing the high-value potential of geographic information spatio-temporal big data, empowering the construction of innovation zones for digital economy development, cultivating the market of

geographic information spatio-temporal big data elements, and promoting the marketization of geographic information data elements.

3.4 The Applications of Geographic Information Data Elements Circulation

3.4.1 The National Platform for Common Geospatial Information Services (TIANDITU): An open and shared service portal of networked geographic information built by National Geomatics Center of China (NGCC), as shown in Figure 2. It integrates geographic information public service resources from national, provincial, city (county) surveying and mapping departments at all levels, as well as relevant government departments, enterprises and institutions, social organizations, and the public to provide authoritative, standard, and unified online comprehensive geographic information services to all kinds of users.



Figure 2. The National Platform for Common Geospatial Information Services (TIANDITU) portal website. (<https://www.tianditu.gov.cn/>)

3.4.2 Real 3D China: As a real, three-dimensional, temporal and spatial information reflecting human production, life and ecological space, real 3D is an important new national infrastructure, as shown in Figure 3. It can realize the real-time correlation and interconnection between digital space and real space through "human-machine compatibility, perception of things, ubiquitous services", and provide a unified spatial positioning framework and analysis basis for Digital China. It is an important strategic data resource and production factor for digital government and digital economy.



Figure 3. Real 3D China. (Schematic diagram)

3.4.3 The National Natural Resources and Geospatial Basic Information Database:

One of the four basic information databases of national e-government determined by the state. It has formed a basic and strategic geospatial information resource database with standardization, scale and sustainable renewal. A nationwide network service system for geospatial information sharing and exchange, an information resource directory service system, and a mode for large-scale and rapid integration and sharing of multi-source geospatial information have been established. A government information sharing service support system composed of one data master center and 11 data sub-centers was formed. A work system that combines military and civil forces and cross-departmental collaboration and a mechanism for sharing natural resources and geospatial information have been established.

4. Design and Construction of New Spatial Data Infrastructure (NSDI) Supporting the Circulation of Geographic Information Data Elements

The first task of Digital China is to build the new spatial data infrastructure and spatial data resource system. The New Spatial Data Infrastructure (NSDI) is a kind of national Spatial Data Infrastructure (SDI), which can adapt to the characteristics of data elements, promote data circulation and utilization, and play the value of data utility. NSDI supports the whole process of geographic information data aggregation, processing, circulation, application, operation, and security.

4.1 Design of NSDI

The design of the New Spatial Data Infrastructure (NSDI) should contain the cloud data center environment, website portal, and circulation subjects.

4.1.1 The Cloud Data Center Environment of NSDI: Mainly includes cloud computing, storage, security and network resources and corresponding technical services to continuously ensure the safe and stable operation and efficient service support of NSDI, as is shown in Table 1.

Systems	Contents	Core performance metrics
Cloud Hosting Services	Application ECS; Data Storage ECS; Configure OS; Support x86 and ARM architecture	CPU ≥ 8 cores, memory ≥ 16 GB, system disk ≥ 300 GB SSD storage. Among them, the CPU frequency is ≥ 3.0 GHz, the maximum internal network bandwidth is ≥ 42 Gbps, the maximum internal network packet sending and receiving is ≥ 12 million PPS, the maximum throughput of SSD is ≥ 350 MB/s, and the access delay is ≤ 1 ms
Cloud Network Services	Elastic Scaling; BGP Multi-line Access; Gateway Service	Full dynamic BGP network bandwidth; Support HTTP,HTTPS,TCP,UDP load balancing, support tens of millions of concurrent connections
Cloud Databases	PostgreSQL; MySQL	CPU ≥ 32 cores, memory ≥ 128 GB, total storage

		space \geq 30TB SSD storage, SSD read and write speed up to 250MB/S, read and write latency \leq 1ms, cluster concurrent QPS \geq 30000 times /S, query response time is not more than 100ms
Domain Name Resolution Services	First and second level domain name resolution service	DNS protection traffic \geq 80G; DNS protection QPS \geq 80WQ/s, A record load balancing \geq 90
Cloud Monitoring Services	Operation and maintenance monitoring	Monitoring and management of all resources in public cloud
Network Security Services	Web Application Firewall; Situation Awareness; Container Security; Host Security Service; Database Audit Service; Vulnerability Scanning Service; SSL Certificates; Fortress Machine	Protection bandwidth \geq 1GB; Supports multi-dimensional CC defences; Provides a variety of threat detection and analysis services, including threat analysis, alarm setting, host vulnerability, website vulnerability, baseline inspection, security log management and other functions; Container images and operational security; Malware detection, web tamper proof; Provide behavior/data/performance anomaly monitoring; Web site scanning and host scanning; Business data transmission security; Real-time authorization from the authorizer is required to access the resource
Cloud Search Services	Query data; Vector Retrieval	8 cores 64GB, 3 nodes, high IO, 1000GB storage
Big Data Services	Data Migration Service; Data Warehouse Services; Data access service	The number of data access service partitions is 15, the size of the number of messages is 1KB, and the number of messages written is 20000 per second; Batch data acquisition; CPU \geq 16 cores, memory \geq 128GB, number of nodes \geq 3, total data storage empty 4000GB SSD storage

Table 1. The Table of Geospatial Data Infrastructure Components.

4.1.2 The Website Portal of NSDI: Mainly includes home page, about us, data services, news and identity authentication, as is shown in Table 2.

Subject of web portal	Contents
Home Page	The home page of the platform portal
About Us	NSDI introduction, Terms of Service, Copyright statement, Contact us, feedback
Geographic Information Data Elements Services	Data Sharing Exchange, Data development and utilization, Data circulation. For example, map service, catalogue service for geographic information, development resources and APIs, thematic data for natural resources, geographic information spatio-temporal big data trading area, standards and specification, tools and applications...
News and Information	Highlights, News Feed
Unified Identity Authentication	User registration, User login

Table 2. The Table of Cloud Data Center Environment Systems.

4.1.3 The Circulation Subject of NSDI: Mainly includes the products generated by the acquisition, processing and analysis of raw data through surveying and mapping means, and the specific forms include but are not limited to API/SDK interfaces, offline data sets, data reports, data system deployment and other data services. The underlying transaction forms are approval offering, transaction offering and sharing offering, as is shown in Table 3.

Subject of transaction	Contents
Basic achievement products	The use of financial funds, through the formation of surveying and mapping data, information, maps and related basic results of information, such as the national unified surveying and mapping standards and surveying and mapping system of the data, maps, satellite geodesy, gravity survey, basic aerial photography to obtain data, image information; Basic geographic information remote sensing data obtained from Earth observation by remote sensing satellites and other space vehicles, national basic scale topographic maps, image maps and their digital products, new geographic information results of surveying and mapping, etc.
Comprehensive achievement products	In order to serve the macroeconomic and social development, carry out research results such as data, information, graphs, statements and reports formed according to statistical categories, such as government information disclosure of bulletins, announcements and circulars, media publicity reports of newspapers

	and websites, and academic research papers published in comprehensive publications such as books and journals.
Special Outcome Products	Regional data, information, maps, statements, reports and other results formed in order to meet the needs of a certain aspect and carry out special investigation and monitoring research.
Special achievement products	Specific data, information, graphs, reports, reports and other results required to support a specific business for the service, such as single planning, etc.
Exclusive result products	Patent data, information, graphs, statements, reports and other results of independent research, integration and development by enterprises or individuals, such as APP applications, intelligent products, equipment and equipment manufacturing results, etc.
Integrated result products	Integrate all kinds of data, information, maps, reports, reports and other results required in the initiation, feasibility study and construction plan of major projects and major projects, such as models established in construction projects, maps drawn, and surveying and mapping benchmarks.
Other products	Other geographic information spatio-temporal big data achievement products generated according to law.

Table 3. The Table of Subject of Geographic Information Transaction.

4.2 Construction of NSDI

4.2.1 Network of Geographic Information Data: Mainly relies on the construction of national e-government network and Internet.

National e-government network is composed of internal network and external network. The two networks are physically isolated, and the external network is logically isolated from the Internet. The government Intranet is mainly the office network of the government departments above the deputy provincial level, which is physically isolated from the office network of the government departments below the deputy provincial level. Government external network is the business network of the government, which mainly runs the professional service business of the government department for the society and the business that does not need to run on the network. Geographic information public data rely on the Internet.

4.2.2 Geographic Information Data Spaces: The International Data Space (IDS) is a framework that leverages existing standards and technologies, as well as widely accepted governance models in the data economy, to facilitate standardized data exchange and data linking in a secure and trusted business ecosystem. The Trusted Data Space (TDS) is the technical guarantee of the data elements circulation system. By building the distributed critical data infrastructure of data collection, sharing, circulation and application on the existing information network, the systematic technical arrangement can ensure the performance and maintenance of the signed data circulation agreement. It solves the security and trust problems among data element providers, users, service providers, regulators and other subjects.

Geographic information data spaces is through the core module trusted data connection platform, to create a secure data circulation pipeline, to achieve "the original data does not go out of the domain, the data available is not visible", as is shown in Table 4.

Key Parts	Contents
Data Provider Engine (DPE)	The Data Provider Engine is deployed in the data provider domain, the open space of the provider is isolated from the core data area, and the open resources such as data warehouse, relational database, API, and data set files are flexibly configured.
Data Consumer Engine (DCE)	The Data Consumer Engine is deployed in the data consumer domain: the flow batch security sandbox of the consumer is integrated, and the data inside and outside the domain is seamlessly connected to calculate and use under the data usage control strategy.
Secure Pipeline (SP)	Between DPE and DCE point-to-point, a secure data circulation pipeline is formed: cross-domain data sovereignty signaling protocol stack, which realizes device mutual trust, device authorization, application negotiation authorization, application session control, usage control, data push and pull.
Micro-isolation Sandbox (MS)	The data was processed in real time in the micro-isolation sandbox of the DPE to be opened, to ensure that the data obtained from the initial processing could retain its business characteristics, and to ensure that the original data could not be reversed.
Joint Computation (JC)	In the micro-isolation sandbox of the DCE joint computing space, the relevant data of all parties and the preliminary processed common data authorized from the DPE through the cross-domain data usage control security pipeline are fused, and the controlled joint computing is carried out in the form of cross-domain virtual tables to form higher-order data and model results that meet requirements.

Table 4. The Table of Geographic Information Data Spaces Components.

5. Other Aspects of Geographic Information Data Circulation that Need Attention

5.1 Classification and Grading of Geographic Information Data

Geographic information data are divided into basic geographic information data, remote sensing image data and thematic geographic information data. Basic geographic information data refers to the geographic information data as the basis of the unified spatial positioning framework and spatial analysis, including positioning basic data, place name address data, terrain and geomorphology data, etc. Remote sensing image data refers to the image data of the earth's surface obtained from the ground, air and space by using various sensors on the remote sensing platform, including original remote sensing image data and remote sensing image product data. Thematic geographic information data refers to the geographic information data that takes basic geographic information data as the spatial positioning framework and spatial analysis basis, emphasizes one or several elements in natural geography and social economic and cultural phenomena, and focuses on the expression of a certain theme content, including natural resources, ecological environment, disasters and disasters, economy and society, infrastructure and other thematic geographic information data.

According to the importance, accuracy, scale, depth, current situation and security risk of geographic information data, it is divided into three grades: general, important and core, as is shown in Table 4.

Grade	Contents
Geographic information core data	The core data of geographic information refers to the important data of geographic information that has a high coverage or reaches a high precision, a large scale and a certain depth for the field, group and region. Once illegally used or shared, it may directly endanger political security.
Geographic information important data	Geographic information important data refers to geographic information data in specific fields, specific groups, specific regions or reaching a certain accuracy and scale. Once leaked, tampered or damaged, it may directly endanger national security, economic operation, social stability, public health and safety.
Geographic information general data	Geographic information general data refers to other geographic information data other than geographic information core data and geographic information important data.

Table 4. The Grades of Geographic Information Data.

5.2 Geographic Information Data Right Confirmation

Data right confirmation is to empower data processors to enjoy corresponding legal control means over data, so as to have the effect of excluding other people's infringement on data to a certain extent or scope. The characteristics of data are replicable, non-competitive, non-exclusive and non-exhaustible, which makes the acquisition and use of data difficult to be blocked by physical means, and must rely on legal means such as data right

confirmation, otherwise it is difficult to protect the data rights and interests of relevant subjects. In the whole life cycle of data from generation to extinction, there are mainly four types of roles: data owner, data producer, data consumer and data manager. And right confirmation is the process of clearly defining these four types of roles for specific data assets. That is, different data assets may have different owners, producers, consumers, and managers.

5.3 Geographic Information Data Elements Security

Generally speaking, there is still a big gap between the openness and utilization level of public data in our country and the expectations of all sectors of society. It is widely reported that there are many obstacles and concerns about data sharing and opening up, the scale and quality of data supply are not enough, and the channels and methods of resource utilization are not abundant and convenient. The main reason is the lack of motivation of data holders and the "reluctance to open" problem. The large-scale circulation of data leads to the expansion of security management boundaries, and there is a problem of "dare not open". It is difficult to balance the efficiency, security and cost of data circulation, which needs professional technical ability to support. Some units have the problem of "will not open". To solve these problems, we should not only continue to increase the work of sharing and opening up, but also summarize the practical experience of authorized operation, adopt security methods to develop high-value data, and solve the contradiction between security protection and circulation and utilization. The main technical means include the digital watermark, blockchain, federated learning, secure multi-party computation, homomorphic encryption, confidential computing and so on.

6. Challenges and Opportunities of Geographic Information Data Circulation

In order to vigorously promote the development and utilization of public data resources and accelerate the reform of data marketization configuration, it is necessary to focus on the following aspects: First, implement the system of separation of property rights, clarify the compliance policies and management requirements for the authorized operation of public data, clarify the rights and obligations of data supply, use and management, explore the price formation mechanism of public data products and services, and make public data "available". Second, accelerate the construction of secure and trusted data infrastructure, develop data space and high-speed data network, promote the application of privacy computing and blockchain technology such as anonymization, federated learning, and multi-party secure computing, enhance the trusted, controllable, and measurable ability of data utilization, and let public data "flow". The third is to implement the data factor x action plan for the pain points of the industry development, form a number of typical application scenarios that serve economic and social development, give full play to the multiplier effect of public data, and let public data "be used well".

References

Chen J, Wang Y H, Wu H, Liu W Z. Basic issues and development directions of high-quality development empowered by spatio-temporal information. *Journal of Spatio-temporal Information*, 30(1): 1-11. doi: 10.20117/j.jsti.202301001.

Harvey J.M., Han J., 2009. Geographic Data Mining and Knowledge Discovery. London: CRC Press.

Jiang W B, Guo W H, 2022. Construction and application of big data on natural resources and territorial space. *Natural Resources Informatization*, 41(5), 1-6.

Li Deren, Ma J., Shao Z. F., 2015: On spatiotemporal big data and its applications. *Satellite Application*, 09:7-11.

Li Heng, Huang Wei, Zheng Xinyan et al., 2022: Application and Platform Design of Spatiotemporal Data Opening and Sharing. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, Volume 43, Issue B4-2022, Pages 369-376, May 30, 2022.

Li Heng, Huang Wei, Zhao Wenhao et al., 2023: Research on Geographic Information Data Circulation Supports the Construction of Digital China. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, Volume 48, Issue 1/W2-2023, Pages 97-104, December 14, 2023.

The CPC Central Committee and The State Council, Issued the Overall Layout Plan for the Construction of Digital China. www.gov.cn/xinwen/2023-02/27/content_5743484.htm (21 May 2024)

The CPC Central Committee and The State Council, On Building a Data basic System to better play the role of data elements. (Data 20) www.gov.cn/zhengce/2022-12/19/content_5732695.htm (21 May 2024)

Wang Jiayao, Wu F., Guo J. Z. et al., 2017: Challenges and opportunities of spatio-temporal big data. *Science of Surveying and Mapping*, 42(07):1-7. <https://doi.org/10.16251/j.cnki.1009-2307.2017.07.001>.