A Web-Based Remote Sensing Data Processing and Production System With the Unified Integration of Multi-Disciplinary Data and Models

HONGYI LI, ZHENG ZHANG, AND PING TANG
Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China

Corresponding author: Zheng Zhang (zhangzheng@aircas.ac.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2016YFB0501501, in part by the Special Fund of China Seismic Experiment Site under Grant 2018CSES0201, and in part by the National Natural Science Foundation of China under Grant 41701399.

ABSTRACT
The development of Earth observation and navigation technology has enriched the methods and means of remote sensing data acquisition. With the increase in the data sources, data volume and data velocity, multi-disciplinary studies can use remote sensing data to solve problems in their individual fields, which, however, increases the heterogeneity of multi-source data and the complexity of multi-domain models. In this paper, we propose a Web-based framework of integrated data and models that allows the horizontal extension of the corresponding systems and effectively addresses the heterogeneity of the integration of multi-source data services and the complexity of multi-disciplinary model services. The interoperability of multi-disciplinary models becomes feasible with a unified interface design and Web service construction for interdisciplinary models. Moreover, a global unified geographic location framework and an active data service based on GeoHash optimized storage are designed to enable fast data retrieval and on-demand active service, which effectively solves the problem of the multi-source heterogeneity of remote sensing data. Ultimately, a global multi-source remote sensing data processing and production system with integrated data and models is implemented.

INDEX TERMS
Multi-source data, heterogeneous data, remote sensing system, model integration, data integration, model service.

I. INTRODUCTION
The profusion of Earth observation and navigation techniques in recent years has greatly enhanced the acquisition capability of remote sensing data. The increase in data sources, data volume and data velocity has expanded the application scenarios of remote sensing applications. On the one hand, it has extended the spatial scopes of traditional remote sensing applications from small regional scopes to the global scope, which could even be further extended to the three-dimensional range of deep-space and deep-Earth locations. On the other hand, it has expanded the disciplinary scope of applications and re-enabled applications that used to be constrained by their data acquisition abilities, data openness, technological levels, etc. In addition, with the development of interdisciplinary studies, it is becoming more necessary to use data and models from other disciplines to solve problems in individual fields.

The improvement of data acquisition platforms, data processing technology and computational capabilities has opened new opportunities for global and multi-disciplinary remote sensing applications. The research horizons of scientists have been extended to the global scale, yielding a number of global remote sensing applications, such as global land cover classification [1], [2], global water resource distribution [3], global crop yield estimation [4], global crop pest monitoring [5], and moderate resolution imaging spectroradiometer (MODIS) and Fengyun (FY) products based on a single sensor source. To support the production of global remote sensing products, some integrated and automated production systems for specialized products have emerged, for example, GLASS [6] and MuSyQ [7], [8]. Although GLASS
integrates five different types of remote sensing products, and MuSyQ integrates 40 different types of remote sensing products, neither of them employs a service-oriented data integration architecture or a service-oriented production model integration architecture. This is not conducive to continuous model integration or the large-scale data expansion of production systems. Another well-established production system is NASA’s Earth Observing System Data and Information System (EOSDIS) and its Distributed Active Archive Centers (DAACs) [9], where each DAAC is responsible for managing a particular discipline of Earth Observing System (EOS) data. However, this subdivision is not suitable for multi-source synthesis production and would cause heavy data transmission pressure when working with multi-disciplinary products. In addition, a DAAC manages data with file indices and file systems, which is not as flexible as service-oriented data management approaches when integrating new models or new data.

The expansion of the depth and breadth of remote sensing applications brings new opportunities as well as challenges for the remote sensing community in constructing large-scale remote sensing applications. Faced with a large number of heterogeneous data and models, the building of a comprehensive production platform that integrates data acquisition, data processing, and model orchestration poses three main difficulties and challenges, as follows:

1) The difficulty of systematically managing multi-source heterogeneous data increases with the amount of data. Taking remote sensing satellite data as an example, 514 Earth observation satellites have been launched since the first launch, in December 2011 [10]. In addition, increasing numbers of Earth observation programmes have been executed in recent years, such as the Copernicus Plan of the European Union and the China High Resolution Earth Observation Plan, resulting in a fast increase in the amount of data. For example, China Remote Sensing Satellite Ground Stations (RSGSs) have received, processed and archived more than 3 million scenes of Landsat TM/ETM+/OLI data, requiring more than 200 TB of data storage space. Such increases in the amount of data pose great challenges for data storage and data access [11], [12]. Moreover, the diversity of data structures, ranging from single remote sensing raster data to the co-existence of multiple data structures [13], [14], including raster data, point data, and structured and semi-structured data, makes it an urgent task to build a unified geographic framework for multi-source heterogeneous data to achieve unified data organization and retrieval.

2) The complexity of model integration is increasing. The algorithms of multi-source remote sensing products usually require inputs with varying numbers of parameters or varying data formats. These algorithms and their inputs can even be hierarchically nested for the calculation of the final products, which causes difficulties for the automation of production and the integration of algorithm models. In this situation, the integration of models from a single traditional remote sensing domain in multiple interdisciplinary domains becomes even more complex.

3) The pressure on computation and storage is surging. Due to the integration of an increasing number of data and models within a single platform and the global scale of production tasks, the demands on computation and storage capacity have increased sharply. An efficient design for data and model integration is becoming a necessity.

In this paper, we aim to build a unified system to enable the management of multi-source heterogeneous data and the integration of multi-domain models. The system serves as the main platform to support the production of global spatial information products in the context of the Earth Big Data Science Project, a Strategic Priority Research Program of the Chinese Academy of Sciences (referred to as the Priority Project). In the project, researchers from various disciplines, for example, environmental studies, resources, ecology, and biology, propose their own spatial information products and the corresponding algorithms, while the system serves as a one-stop platform for the entire life cycle of these products, including source data indexing, data pre-processing, product calculation, and product sharing. In the current phase, the 23 kinds of spatial information products listed in Table 1 will be integrated into our system. The design of the system also has to be flexible enough to enable the integration of other future products.

II. BACKGROUND AND ADVANCES

A. DATA INTEGRATION

In the Priority Project, the integrated management of remote sensing data mainly involves two data types: remote sensing raster data and metadata (point data and structured and semi-structured data). Remote sensing raster data constitute the majority of the data, and most of the data storage and representation are related to the map projection. Different map projections are used in different application scenarios. Therefore, it is very important to solve the problem of the cooperative use of multi-scale remote sensing data efficiently and accurately. Metadata are to some extent a supplement to remote sensing raster data in the remote sensing field, and can be viewed as an independent object.

Raster data from different satellite sensors are often organized or released in different formats; for example, MODIS data are published in HDF4/HDF5 format; Landsat TM/ETM+/OLI data are usually published in GeoTIFF format; and some other data are organized in RAW format. Accordingly, metadata from different sensors also have distinct formats or attributes. Data heterogeneity in this article refers to differences in data and metadata formats, and in the corresponding data access functions. The proposed system operates on data and metadata with different formats or different access functions through a unified set of abstract interfaces. Given concrete implementations of the interfaces,
TABLE 1. List of spatial information products to be integrated.

| No. | Spatial Information Product Name                                      |
|-----|---------------------------------------------------------------------|
| 1   | global 500 metre forest coverage                                   |
| 2   | global 500 metre cropland coverage                                 |
| 3   | global 500 metre annual coverage percentage of water bodies         |
| 4   | global 500 metre annual coverage percentage of snow and ice         |
| 5   | global 500 metre true colour images                                 |
| 6   | global 500 metre normalized difference vegetation index             |
| 7   | global 500 metre leaf area index                                    |
| 8   | global 500 metre fraction of absorbed photosynthetically active radiation |
| 9   | global 500 metre land surface phenology                             |
| 10  | global 1 km surface evapotranspiration                              |
| 11  | 0.05° aerosol optical depth                                         |
| 12  | global 100 km mapping-xCO2/xCH4                                    |
| 13  | 30 metre surface reflectance                                        |
| 14  | global 30 metre land cover types                                   |
| 15  | 30 metre land cover types in China                                  |
| 16  | 2-30 metre land use types                                           |
| 17  | 10-30 metre fraction of absorbed photosynthetically active radiation |
| 18  | 10-30 metre leaf area index                                         |
| 19  | 10-30 metre leaf chlorophyll content                                |
| 20  | 10-30 metre gross primary productivity                              |
| 21  | 10-30 metre trophic state index of inland water bodies              |
| 22  | 10-30 metre transparency of inland water bodies                    |
| 23  | 10-30 metre chromophoric dissolved organic matter of inland water bodies |

other types of data can also be applied seamlessly in this system.

The standards ISO19115 and ISO19115-2 have been established to manage heterogeneous data, especially geospatial data. ISO19115 defines the attributes—such as the identification, quality, space, time, content, spatial reference system and distribution—of geographic information and services. ISO 19115-2 is an extension of image and grid data that defines information such as the acquisition and band description of satellite remote sensing data.

To enable the collaborative use of multi-source and multi-scale remote sensing raster data, standard specifications for remote sensing data subdivision have been established, such as GB/T 12409-2009 and Open Geospatial Consortium (OGC) discrete global grid systems (DGGSs). GB/T 12409-2009 defines data subdivision with Cartesian plane coordinates, longitude coordinates, and latitude coordinates. OGC DGGSs integrate grid, vector and point data to organize, use and visualize data without the need for projection. The definition of data subdivision has gone beyond the limitations of projection.

GeoHash [15] and GeoSOT [16] are the prominent representatives in the areas of metadata management, the subdivision organization of remote sensing raster data and the integrated association management of data. The logically hierarchical recursive subdivision of the Earth sphere has the characteristics of global uniqueness, hierarchical recursion and linear one-dimensionality, but it avoids the physical dissection of remote sensing raster data entities. GeoHash was initially used for location-based services (LBSs) and point of interest (POI) services, and can also be used for the retrieval of remote sensing grid surface data [17]. GeoSOT has a similar principle to GeoHash, and it is now used for applications in three-dimensional circles and in time scales.

Regarding active data services, with the in-depth application of OGC WMS/WMTS and other specifications, the remote sensing raster data to be displayed in the field of remote sensing raster image display, such as World Wind [18], Google Earth [19], and Skyline Globe are all active on-demand services. The client is responsible for requesting data from the network using the screen area and the display level as the request parameters.

B. MODEL INTEGRATION

A model is an abstract representation of the real world that can help better understand the past, the present and the future of natural systems. In the Priority Project, two main kinds of remote sensing scientific models are involved, i.e., the vegetation model and the surface radiation model. An increasing number of researchers are studying methods of inter-model integration, and they have proposed some solutions [20]. Using Web service interfaces [21], [22], the decoupling of the front-end application and back-end algorithm components is helpful in building a large-scale distributed system and an integrated multi-domain model. Based on the model of open standards [23], the use of the web processing service (WPS) protocol of OGC as the access interaction standard for the front-end and back-end models is conducive to the reuse of the back-end model. Using a component architecture [24], the components are developed in a componentization and micro-service mode. With the open-model interface [25], multi-domain model integration is achieved by providing standardized interfaces for different domain models.

The integration of multi-disciplinary domain models, which combines all models from top to bottom, makes it difficult to specify the time, space, events, attributes, interface requirements and internal data structure of the models. This model integration method is typically used within a single field. The integration of multi-disciplinary modules usually involves several disciplines, which makes it a complex system, and the interoperability and connectivity among domain models need to be considered.

III. METHODOLOGIES

To achieve the integration of multi-source heterogeneous data organization and multi-domain models, the main idea of the system design is to package the multi-source heterogeneous data and multi-type models involved in the Priority Project into services by utilizing the distributed integration capability of the Web and the distributed characteristics of the existing
spatial information infrastructure. The data service internally considers the problem of multi-source heterogeneity, and the model service internally considers the problem of the complexity of multi-domain models. The back-end data and model processing services are geographically decentralized, so they can make full use of distributed computing resources for global and regional data computation, and the services are published on the Web, where the front-end constructs a unified portal to integrate the display data and models.

In contrast to the traditional passive data access paradigm, where each individual algorithm model has to implement access to its own data, an active data service provides an independent data service that implements access to different types of data actively. With an active data service, each algorithm model only needs to invoke existing services according to its own requirements, without considering low-level physical storage or data structures. The active data service decouples the data access from the algorithm model and provides a unified set of demand-driven data access interfaces, which is beneficial for the simplification of data access, the integration of algorithm models, and the expansion of the system.

In addition, models directly obtain data from networks instead of disks. Since networks provide more convenient and scalable access to data—and in cases where advanced high-speed equipment is employed, networks can have a faster IO speed than hard disks—the proposed active data service can further improve the accessibility of data and the computational efficiency of models.

The advantage of the model service is that each model is released as a service on the Web, and the invoker does not need the specific address of the model. By calling the published uniform resource locator (URL) of the model, the model can easily be integrated into the unified Portal framework.

The difficulty of model integration mainly lies in the extensibility of the data streams that support the calculation of models. Taking remote sensing data as an example, each kind of input data has to pass through different pre-processing procedures, and each has a unique data format that may involve distinct bands, description fields, processing levels, and quality standards. If the current system cannot provide full access to the new data types required by new models, the production workflow cannot be assembled, which means the new models cannot be integrated. The active data service aims to solve the aforementioned problem and to provide unified data access that mitigates the differences among possible data types. With the active data service, a new model no longer needs to worry about the compatibility of its input data.

A. SYSTEM ARCHITECTURE

Using the infinite horizontal expansion ability of the Web, data and models are encapsulated as services that mitigate the geographical decentralization of data and model services and integrate multiple types of data and model services on a unified web portal to provide services for users. The overall structure of the system is divided into four layers, as shown in Fig. 1: the basic services, networks, web portal and user. The main functions are distributed in the layers of basic services and the web portal.

The basic services layer is the system server, which encapsulates all data and model services. The data services include remote sensing raster data services and metadata services.
The model services include the vegetation model services and the radiation model services.

The web portal layer is the system portal, which is a window for the display of the back-end service. All data, models and services are displayed to users through the web portal.

**B. MODEL SERVICE**

The integration of the model service includes three steps, as shown in Fig. 2: the production architecture of the products, the model definition & scientific workflow integration, and web interface encapsulation.

Production architecture of the products: the global basic spatial information product architecture is constructed according to the type of satellite platform, sensor type, time resolution, spatial resolution, and sensor band settings as well as the production relationship among the different information products.

Model definition & scientific workflow integration: the time, space, events, attributes, interfaces and other requirements of the model are defined abstractly and formally, the basic spatial information model is arranged with the unified modelling language (UML) in the Scientific workflow, and the automatic workflow is constructed through the combination of the models.

Web Interface Encapsulation: this encapsulates the orchestrated workflow into a Web interface, injects it into the URL, and publishes it on the Web.

Taking the vegetation model as an example, the overall architecture of the product production is established according to the production hierarchy relationship among the models, as shown in Fig. 3. The nine vegetation models shown in blue include the vegetation index, vegetation fine classification, leaf area index, surface coverage classification, phenology, vegetation coverage, chlorophyll content, fraction of photosynthetically active radiation, and net primary productivity of the vegetation. The models shown in green, such as the normalized reflectance, surface albedo and photosynthetically active radiation, can be considered models in other fields. In designing the production model, the interface relationship of the vegetation model is uniformly defined at first, and it is followed by the consideration of the interface, interoperability and interconnections with models in other domains.

The definition of the levels of the model: the normalized reflectance is defined as the starting point and is assigned level 0. The vegetation index, vegetation fine classification and surface cover classification are included in level 1.
Level 2 includes the phenological data, vegetation coverage, chlorophyll content and leaf area index. Level 3 includes only the photosynthetically active radiation. The last level (level 4) includes the production level of the net primary productivity. Combined with the interface definition of the model, it will be organized into a model production framework.

The aim of the model definition is to formalize the rules, describe them in a computer language, and build models that can be recognized and processed by computers. The modelling is implemented in the XML language. For each input parameter of the model, seven basic metadata are defined: the data subdivision mode, the parameter order, whether the product is at level 0 or not, the spatial resolution, the temporal resolution, the time interval, and the time interval alignment. The output parameters of each model are summarized by four metadata: the data partition, spatial resolution, temporal resolution, and product level.

C. ACTIVE DATA SERVICE

To implement the unified external service of multi-source remote sensing raster data and metadata, the service framework of data as a service is shown in Fig. 4. It is divided into five layers; from bottom to top, these are the data layer, the database (DB) layer, the middleware layer, the server container layer, and the web interface layer.

The Data layer is an entity that contains all kinds of data, primarily including remote sensing raster data and metadata. The DB layer is the storage method for all kinds of data. The remote sensing raster data are stored in a spatial database or a relational database with a file system. The metadata are stored in a spatial database or relational database. The structured and semi-structured attribute data are stored in a relational database or NoSQL database. A GeoHash code field is added to represent the spatial location of data.

The middleware layer contains the middleware and mode of data access. Mature middleware can shield the format and structural complexity of certain data. The relational database uses the Hibernate middleware to simplify data access. Remote sensing raster data uses Geospatial Data Abstraction Library (GDAL) middleware for data access. The GDAL integrates various kinds of format parsing for remote sensing raster data, which reduces the costs of reading and writing data.

The data access mode is the interface for accessing different types of data. Remote sensing raster data access follows the storage access mode of the OGC WMS/WMTS protocol. Metadata follows the storage access mode of Sql/Hql. The server container layer is a container of data services. The container provides web services for all data access. The server container is responsible for receiving and forwarding.
data requests and feedback, and Tomcat is responsible for the release of data services. Nginx is responsible for routing data requests to the Tomcat server where the service is located.

The use of Tomcat and Nginx can effectively alleviate the density problem of remote sensing models. Remote sensing data are distributed to different servers, data services are published independently with Tomcat, and Nginx is routed through different Tomcat servers, thus avoiding the IO bottleneck of a single data server.

The web interface layer provides the standard for protocols with which the external services of models and data have to comply.

A unified geographic reference is defined, and with the development of geographic information technologies, geospatial data are multi-source, heterogeneous and multi-temporal. Effective data management and efficient query analysis is one of the key problems to be solved in geographic information systems. However, in reality, the inconsistency of time and space benchmarks often makes it challenging to make full use of these massive data, which seriously limits the effective use of the resources.

A large number of multi-source heterogeneous remote sensing data are involved in the Priority Project and to collaboratively manage the point and polygon remote sensing data of our system, our system is built upon GeoHash coding to correlate the multi-source heterogeneous remote sensing data. To further accelerate the query and retrieval efficiency of multi-source heterogeneous remote sensing data, GeoHash prefix coding is used to redundantly store the geographic location fields represented by GeoHash coding in the database. The system stores 8 32-bit GeoHash codes with coding lengths of 1-7 and 12. The precision of the GeoHash coding with a coding length of 12 is on the scale of centimetres, which meets the requirement of the query precision. In a query, the GeoHash field is automatically determined according to the scale of the geographic location in the query condition. For example, a query for a field of code length 6 is 32 times more efficient than one for a field of code length 7.

Raster data organization is required, and data as a service designs a unified service framework for the access of multi-source heterogeneous remote sensing data. In this section, we introduce how to store and organize multi-source heterogeneous data in order to implement active data services.

Different types of remote sensing data can use different data organization and management methods according to their different application scenarios. In particular, for the same type of data, the storage organization is also different when the application scenario varies. For example, the typical application scenarios of sensing raster data include visual display and quantitative processing, which correspond to two different forms of data organization: (1) Remote sensing...
raster data visualization is organized on the basis of the hierarchically recursive subdivision of patches after sphere flattening (for a rectangular coordinate system) as the basic display unit. (2) The quantitative processing of remote sensing raster data is basically designed for the original scene separation and large-scale single-layer subdivision of remote sensing data.

In our system, we design three kinds of data organization and transformation rules for multi-source and heterogeneous remote sensing data. First, the original remote sensing data, which are connected into the system, are not directly subdivided because there is no demand for on-line display. Second, the advanced products produced in the system need to be displayed on-line, which requires hierarchically recursive partitioning. The active data service takes the hierarchically recursive partitioning size as the criterion for data acquisition, and each request at least has a partitioning surface. Third, all data products that need to be used collaboratively are subdivided according to single-layer partitioning rules, and the data are acquired on demand according to the size requested by users when the data are actively served.

IV. SYSTEM IMPLEMENTATION AND RESULTS
The multi-source heterogeneous remote sensing satellite data involved in the system primarily include HJ series, Landsat 8, Terra/Aqua, NOAA series, ZY3, Sentinel, GF series, and some sampling point data with longitude and latitude coordinates. The system includes 23 models of the global spatial basic information products, e.g., vegetation and radiation models.

A. SYSTEM IMPLEMENTATION
According to the overall logic of the design of the system, the important modules of the system have been refined and implemented according to the domain model. The encapsulation of the remote sensing models aims to encapsulate the remote sensing spatial information models, to establish a unified interface, and to design a suitable interface for other disciplines. The domain model of the encapsulation is shown in Fig. 5. The remote sensing data organization includes the unified geographic location framework of the multi-source heterogeneous data and the subdivision organization of the remote sensing raster data. The domain model of the unified geographic location framework is shown in Fig. 6. The IWFQpModel has three main interfaces: (1) GetInputPara, an interface for obtaining the input parameters; (2) GetOutputPara, an interface for obtaining the output parameters; and (3) ToWorkFlowScript, an interface for converting models to script fragments. WFQpModelFactory is the model factory class, and the instance object of a model is created according to the model name. WFQpModelInterface is the calling interface, which has two main types. One is responsible for aggregating all the script fragments in the model into executable script files in a discipline domain, and the other is responsible for the adaptation of the QpInputPara and QpOutputPara of the model to other disciplines.
Data organization includes the unified geographic location framework of the multi-source heterogeneous data and the subdivision organization of the remote sensing raster data. The domain model of the unified geographic location framework is shown in Fig. 6. IStoreCatalogue is a geo-referenced interface with two interface functions: (1) ToGeoHashes, which converts geographic locations to redundant GeoHash lists. (2) GetByGeoHash, which is retrieved through GeoHash. In the geographic location framework domain model, all classes with names that include "_" have corresponding satellite sensors (i.e., remote sensing raster data), except the StoreCatalogue_Point class, which is used to store point data. StoreCatalogue_HJ1ACCD1 is refined for the class of sensors, such as StoreCatalogue_HJ. StoreCatalogueFactory is a factory class, which returns the corresponding instance according to the sensor type.

The domain model of remote sensing raster data is shown in Fig. 7. ICoverGridInfo is the interface for partitioning, which has three interface functions. The first function is AssemblyGridFileName, which is the filename for assembling facets. The second function is CalCoverGridInfo, which calculates the list of split faces corresponding to remote sensing raster images. The last function is SplitImage, which is a physical subdivision of remote sensing raster images. CoverGridInfo/WebMercator, CoverGridInfo_LngLat, CoverGridInfo_UTM and CoverGridInfo_Sin are the four subdivision classes, which correspond to WebMercator, longitude and latitude, uniform transverse Mercator and sinusoidal projections, respectively. The above are the single-layer projections of common remote sensing raster images. CoverGridInfoFactory is a partitioning factory class, and it automatically returns the corresponding partitioning instances via different projection types. The CoverGridInfoRecurse class relies on the four subdivision classes to implement a hierarchically recursive partitioning of the four projections.

The domain model of the unified geographic location framework implements a unified query and retrieval criterion for multi-source heterogeneous data. The subdivision domain model implements a uniform hierarchically recursive partition of multi-source, multi-scale and other projection data. The active data service interface combines these two domain models to achieve fast data retrieval and on-demand data acquisition.

The Web service encapsulation of models and data adopts the J2EE platform. As shown in Fig. 8 (a) and (b), the above models and data interface functions are implemented by
As shown in Fig. 8 (c), models and data in different disciplines are designated to geographically dispersed Tomcat containers. Routing in the Nginx server is configured to achieve the distributed integration and processing of models and data. In particular, in order to ensure the versatility of the front-end applications of the display-oriented remote sensing data, the raster data acquisition follows the OGC interface standard.

B. RESULTS
The proposed system implements a Web-based framework for the integration of models and data. A domain-driven design strategy is employed throughout the implementation of the entire system, which makes it flexible enough to add
new remote sensing spatial information models and new data sources.

The inclusion of a remote sensing spatial information model only requires adding a class that inherits from WFQpModel, without the need to override a function from the base class in most cases. In addition, since the WFQpModelFactory class uses the Reflection mechanism, it can instantiate new remote sensing spatial information models without further coding or modification.

Similarly, the inclusion of a remote sensing data source only inheriting from the StoreCatalogue class, and in most cases, it is not necessary to override the base functions to instantiate the inherited class via the factory class. The implementations of the data subdivision include the four types of common projections that meet the demands of the production of global spatial information product models in the current phase. In subsequent studies, the inclusion of other projection types can be as convenient as the inclusion of the models and data sources.

In addition, due to the separate development modes of the Web server and the client, the increase of model types and data sources can be directly reflected in the Web front-end. There is no need to adjust the code structure of the Web front-end or to add new codes to the Web front-end. Fig. 9 shows a series of thumbnail images of the partial subdivision cells of a global 1KM MODIS mosaic image, partitioned by a data partitioning service. The projection type of the service request is WebMercator. The processing and exhibition of the information model only require visiting the active data service via the URL interface, where the data are accessed on demand (with a specified projection, specified size, and specified band). The application side does not need to be concerned with the underlying details of the data format, the projection type or other related information. Data transparency is achieved by the active data service of the proposed system.

For the production service of the integrated products, users need to visit a URL such as {hostIP}/radi/doWorkInfoProduct via the HttpRequester plug-in, and indicate the type name, time range and spatial range (quadrangular coordinates) of the products in the parameter input fields, as listed in the interface function parameters in Fig. 8 (b). For example, the production of the normalized difference vegetation index (NDVI) requires the following input parameters: the product type is NDVI, the time range is from 2018/04/01 to 2018/04/16, and the spatial range described in longitude/latitude form is (E70.71/N54.52, E136.27/N5.65). The result of the NDVI products in China is shown in Fig. 10.

V. CONCLUSION

Based on a Web application framework, we propose an integrated system of data and models where “everything is a service”. Considering the fact that the spatial data are distributed in multi-centres and that the scale of data storage and processing in the big data era greatly exceeds the capacity of single-centre storage and processing, a service-oriented integration mechanism is designed; i.e., all the data and models are integrated in the form of web services. A model does not need to be concerned with the locations and details of data storage, and an application does not need to be concerned with the specific details of the model and computing hardware. Our study makes full use of the lateral expansion ability of the Web to integrate the dispersed data and models, leading to an effective solution for the problems of large-scale computing and global rapid production.

The domain-model-driven software design simplifies the implementation of the system. For a model integration and active data service interface, once the core domain model is established, a new spatial information model together with the new data type it involves can be integrated into the system smoothly without much modification. The proposed data active service can mitigate the complexity caused by data heterogeneity. There are many categories of remote sensing data with various formats and complicated structures. Different models and applications require different data and distinct data organization formats. For example, raster images are mainly displayed on hierarchically recursive subdivision facets as the basic display unit after sphere flattening (in a rectangular coordinate system). Raster images are processed mainly in divided scene units. We do not distinguish specific application scenarios in this study. By building a unified data dissection and publishing it via the Web, we not only shield all the details of heterogeneous data from upper-level applications but also avoid the necessity of storing multiple data spaces for different applications.

Using the optimization-based unified geographic location framework, the association of heterogeneous data and efficient queries are achieved. The multi-source heterogeneity of remote sensing data makes obtaining correlations among data difficult. The unified GeoHash coding is constructed from the multi-source heterogeneous data to achieve the fast geographic association of the multi-source heterogeneous data. The redundant encoding storage of the spatial data enables data queries with shorter indexing code, and therefore the querying process is further accelerated.

The current processing and production system of global spatial information products V1.0 only implements the back-end Web service interface, which requires HttpRequester to access the data and models. Our future work includes the design and investigation of the on-line visualization of data and models and the display of decentralized models and data services on a unified Web platform so that the user interface of the system can be more friendly and more viable.

REFERENCES

[1] P. Gong, J. Wang, L. Yu, Y. Zhao, L. Liang, Z. Niu, X. Huang, H. Fu, S. Liu, and C. Li, “Finer resolution observation and monitoring of global land cover: First mapping results with Landsat TM and ETM+ data,” Int. J. Remote Sens., vol. 34, pp. 2607–2654, 2013.
[2] J. Chen, Y. Ban, and S. Li, “China: Open access to earth land-cover map,” Nature, vol. 514, p. 434, Oct. 2014.
[3] Z. Niu et al., “Mapping wetland changes in China between 1978 and 2008,” Chin. Sci Bull., vol. 57, no. 22, pp. 2813–2823, 2012.
[4] B. Wu, R. Gommes, M. Zhang, H. Zeng, N. Yan, W. Zou, Y. Zheng, N. Zhang, S. Chang, Q. Xing, and A. van Heijden, “Global crop monitoring: A satellite-based hierarchical approach,” Remote Sens., vol. 7, no. 4, pp. 3907–3933, Apr. 2015.

[5] W. Huang, J. Zhang, Y. Shi, Y. Dong, and L. Liu, “Progress in monitoring and forecasting of crop pests and diseases by remote sensing,” J. Nanjing Univ. Inf. Sci. Technol., vol. 10, no. 1, pp. 30–43, 2018.

[6] X. Zhao, S. Liang, S. Liu, W. Yuan, Z. Xiao, Q. Liu, J. Cheng, X. Zhang, H. Tang, X. Zhang, Q. Liu, G. Zhou, S. Xu, and K. Yu, “The global land surface satellite (GLASS) remote sensing data processing system and products,” Remote Sens., vol. 5, no. 5, pp. 2436–2450, May 2013.

[7] Z. Zhang, P. Tang, H. Li, and Z. Feng, “Refined domain model for multisource data synergized quantitative remote sensing production system,” J. Remote Sens., vol. 20, no. 2, pp. 84–196, 2016.

[8] T. Cui, Y. Wang, R. Sun, C. Qiao, W. Fan, G. Jiang, L. Hao, and L. Zhang, “Estimating vegetation primary production in the Heihe river basin of China with multi-source and multi-scale data,” PloS ONE, vol. 11, no. 4, Apr. 2016, Art. no. e0153971.

[9] X. Lu, C. Cheng, J. Gong, and L. Guan, “Review of data storage and management technologies for massive remote sensing data,” Sci. China Technol. Sci., vol. 54, no. 12, pp. 3220–3232, 2011.

[10] Z. Zhang, Y. Ma, G. Wang, X. Zhang, Y. Peng, T. Long, G. He, and L. Wang, “Processing of Earth observation big data: Challenges and countermeasures,” Chin. Sci. Bull., vol. 60, nos. 5–6, pp. 470–478, Feb. 2015.

[11] J. Yan and L. Wang, “Suitability evaluation for products generation from multisource remote sensing data,” Remote Sens., vol. 8, no. 12, p. 995, Dec. 2016.

[12] Y. Shi, M. Zhan, L. Yin, and Z. Wang, “Research on associated organization and analysis of target-oriented multi-source heterogeneous data,” Bull. Surv. Map., vol. 1, pp. 102–104, Jan. 2015.

[13] P. Yue, J. Gong, and L. Di, “Augmenting geospatial data provenance through metadata tracking in geospatial service chaining,” Comput. Geosci., vol. 36, pp. 270–281, Mar. 2010.

[14] R. Moussalli, M. Srivatsa, and S. Asaad, “Fast and flexible conversion of geohash codes to and from latitude/longitude coordinates,” in Proc. IEEE 23rd Annu. Int. Symp. Field-Program. Custom Comput. Mach., Vancouver, BC, Canada, May 2015, pp. 179–186.

[15] H. Song, C. Q. Cheng, G. L. Pu, F. G. An, and X. Luo, “Global remote sensing data subdivision organization based on GeoSOT,” Acta Geodaetica Cartographica Sinica, vol. 43, no. 8, pp. 869–876, Aug. 2014.

[16] A. Jin, C. Cheng, S. Song, and B. Chen, “Regional query of area data based on Geohash,” Geography Geo-Inf. Sci., vol. 29, no. 5, pp. 31–35, 2013.

[17] D. G. Bell, F. Kuehnel, C. Maxwell, R. Kim, K. Kasraie, T. Gaskins, P. Hogan, and J. Coughlan, “NASA world wind: OpenSource GIS for mission operations;,” in Proc. IEEE Aerosp. Conf., Big Sky, MT USA, Mar. 2007, pp. 1–9.

[18] R. J. Lisle, “Google Earth: A new geological resource,” Geol. Today, vol. 22, no. 1, pp. 29–32, Jan. 2006.

[19] J. Wang, M. Chen, G. Lü, S. Yue, K. Chen, and Y. Wen, “A study on data processing services for the operation of geo-analysis models in the open Web environment,” Earth Space Sci., vol. 5, no. 12, pp. 844–862, Dec. 2018.

[20] J. L. Goodall, B. F. Robinson, and A. M. Castronova, “Modeling water resource systems using a service-oriented computing paradigm,” Environ. Model. Softw., vol. 26, no. 5, pp. 573–582, 2011.

[21] C. Granell, L. Díaz, S. Schade, N. Ostländer, and J. Huerta, “Enhancing integrated environmental modelling by designing resource-oriented interfaces,” Environ. Model. Softw., vol. 39, pp. 229–246, Jan. 2013.

[22] A. M. Castronova, J. L. Goodall, and M. M. Elag, “Models as Web services using the open geospatial consortium (OGC) Web processing service (WPS) standard,” Environ. Model. Softw., vol. 41, pp. 72–83, Mar. 2013.

[23] D. S. Peckham, E. W. H. Hutton, and B. Norris, “A component-based approach to integrated modeling in the geosciences: The design of CSDMS,” Comput. Geosci., vol. 53, pp. 3–12, Apr. 2013.

[24] J. B. Gregersen, P. J. A. Gijsbers, and S. J. P. Westen, “OpenMI: Open modelling interface,” J Hydroinform., vol. 9, no. 3, pp. 175–191, 2007.

HONGYI LI received the B.E. degree in computer science and technology and the M.E. degree in computer software and theory from Ningxia University, in 2007 and 2011, respectively, and the Ph.D. degree in signal and information processing from the Chinese Academy of Sciences, Beijing, China, in 2019. He joined as an Associate Professor with the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, where he is currently an Associate Professor with the Aerospace Information Research Institute. His research interests include image processing and analysis, software system designing for satellite image processing and applications, data mining, and scientific programming.

ZHENG ZHANG received the B.S. degree in spatial information and digital technology from Wuhan University, in 2011, and the Ph.D. degree in signal and information processing from the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China, in 2016. He has also been a Visiting Scholar with LETG UMR CNRS 6554, University of Rennes 2, France, in 2019. He is currently an Associate Professor with the Aerospace Information Research Institute, Chinese Academy of Sciences. His research interests include remote sensing image processing, time series analytics, deep learning, artificial intelligence, and high-performance remote sensing production systems.

PING TANG received the B.S. degree in mathematics from Ningxia University, Yinchuan, China, in 1986, and the M.S. and Ph.D. degrees in mathematics from Beijing Normal University, Beijing, China, in 1993 and 1996, respectively. In 1998, she was a Professor with the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing. She is currently serving as a Team Leader and leading a project of multispectral imagery radiometric and geometric correction for large volume of image data at global scale for higher resolution global land-cover mapping. She has significant experience in managing and designing software systems for satellite image processing and applications. Her research interest includes the use of mathematical theories to development algorithms related to image processing and analysis.

***