

# A Framework of Satellite Observation Data Integration System

Rong Xie

International School of Software  
Wuhan University  
Wuhan, China  
e-mail: xierong@whu.edu.cn

Yawen Liu

School of Remote Sensing and Information Engineering  
Wuhan University  
Wuhan, China  
e-mail: liuyawen70@126.com

Xiangxiang Li

Space Star Technology Co., Ltd  
Beijing, China  
e-mail: li\_322503@126.com

Liang Yu

Gold Mantis School of Architecture&Urban  
Environment  
Suzhou University  
Suzhou, China  
e-mail: yuliang\_163cn@163.com

**Abstract**—Under background of rapid development of economic globalization and network information technology, rapid acquisition, effective processing and efficient use of earth observation data becomes the common requirements all over the world. Facing the challenges of massive geospatial information flow gathered from satellite remote sensing to ground observation, the paper proposes a technical framework of satellite observation data integration system. Common metadata model is established for standardized satellite data integration to solve the issue of complex satellite data directory, which can integrate multi-source and heterogeneous data, and also data from different archiving systems. Through cognitive and knowledge discovery method, satellite knowledge base is built to reflect spatio-temporal changes and their association, and remote sensing big data analysis is developed for massive satellite data. Based on shared knowledge base, collaboration technology of multiple data center can be applied to address data storage and sharing of distributed massive satellite data, so that a distributed high-performance satellite data-cloud platform can be established that allows users access valuable data and information directly from different nodes. The proposed framework can provide an effective solution to distributed storage, data format conversion and interoperability for satellite remote sensing big data.

*Keywords*-heterogeneous data integration; remote sensing big data analysis; satellite data warehouse; satellite data-cloud platform; satellite observation data integration system

## I. INTRODUCTION

Many countries all over the world have launched various types of remote sensing satellites, thus an era for global high-resolution observation is coming [1, 2]. Under background of rapid development of economic globalization and network information technology, rapid acquisition, effective processing and efficient use on earth observation data becomes the common requirements all over the world [3]. In particular, the era of big data brings

new challenges for application research of earth observation data [4, 5]. When we meet massive geospatial information flow gathered from satellite remote sensing to ground observation, one hand, we should handle how to make effective integration between satellite data and ground observation data; on the other hand, we should also handle how to transfer these data into useful information which users can take them directly. It becomes an important means to effective management and application for establishment of satellite observation data integration system.

In the recent years, a series of international standards are proposed for satellite data, such as ISO/TC211, OGC etc. [6, 7]. The purpose is to connect global distributed network of satellite remote sensing database through establishment of common standards and protocols, to ensure compatibility and interoperability of satellite data and services, so that all satellite remote sensing data can be standardizedly shared across network. In China, the project "High-Resolution Earth Observation System" [9] is aiming to enhance capacity of high-resolution observation data acquisition to push construction of spatial information system as well as application of satellite technologies. However, the existing satellite data and standards can not be used well, are difficult to use, and can not be obtained easily, which can not fully meet the requirements of development of an integrated satellite data system [9, 10]. Specifically,

- lack of expression of geolocation, geo-referenced and other semantic information of satellite image data in satellite data model.
- lack of definition of spatial relationship between satellite image and ground grid, which can not use geometric correction and accuracy of ground coordinate conversion.
- lack of satellite data analysis which can not integrate the aspects of data integration, spatial

query and data mining for users to automatically extract useful information.

Satellite remote sensing information has become an indispensable strategic and economic resources. In the future 5-10 years, satellite remote sensing data will make explosive growth and also its applications will present tremendous opportunities for promotion and use [11]. Therefore, it becomes an important trend to establish a complete satellite remote sensing application system, which can integrate multi-source satellite and remote sensing data, satellite remote sensing data among different platforms, and satellite data and ground observation data.

To address these problems, the paper focuses on an overall technical framework of satellite data integration system on the basis our research results and improvements [12, 13]. We propose a model to effectively integration of heterogeneous data from multiple sources and different platforms, such as satellite data, ground observation, and simulation model. On the basis of model, we can develop spatio-temporal data query and analysis for massive satellite data, and can reflect spatio-temporal changes and association of ground parameters, and establish a satellite knowledge base. Eventually, a distributed high performance satellite data-cloud platform can be established that users may directly access system to obtain their valuable information.

## II. SYSTEM FRAMEWORK

The basic scientific issues on the study of earth observation, we think, may include two aspects as follows, that is,

- 1) valid data integration approach which can integrate satellite imagery data with ground observation data;
- 2) cognitive science of spatio-temporal data model, and data conversion theory between earth observation data and ground knowledge.

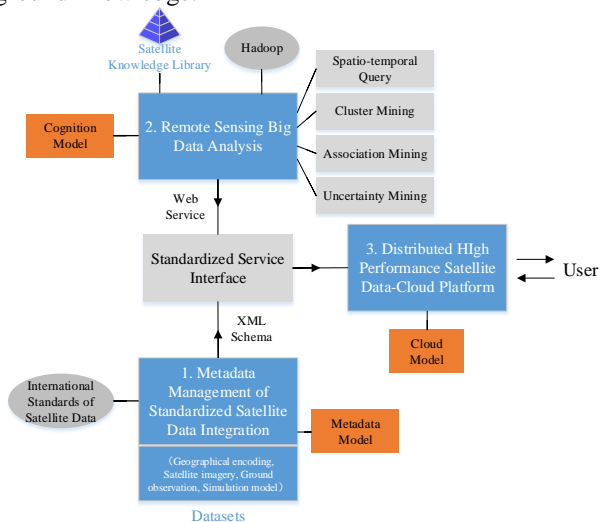


Figure 1. Technical framework of satellite observation data integration system.

As shown in Fig .1, we propose the general technical framework for establishment of satellite observation data integration system. The system includes,

- metadata management of standardized satellite data integration;
- remote sensing big data analysis;

- distributed high performance satellite data-cloud platform.

Here, metadata management is the basis of the whole integration system.

## III. METADATA MANAGEMENT OF STANDARDIZED SATELLITE DATA INTEGRATION

For satellite data integration, it is required to address complexity of satellite data directory firstly. Therefore, a standardized common metadata model is created to integrate effectively various heterogeneous data sources, including satellite remote sensing data, ground observation, and simulation models, and also to integrate different data from archiving systems.

### A. International standards of satellite data

ISO 19115 was recognized for a standard of geographic information metadata by international standard organization in 2003. It defines and describes pattern and service of geographic information, such as representation, range, quality, spatial and temporal patterns, spatial reference and distribution of geographic data, including metadata model, core elements, metadata package, data dictionary etc. In ISO 19115, relevant information relate to image data is data representation and graphic description, positional accuracy of raster data, spatial representation of raster, level descriptions etc. But it does not define metadata elements for all satellite images. ISO 19115: Part II extends metadata of ISO 19115, identifying metadata describing image and raster data, such as data quality, spatial representation, contents and information acquisition etc.

Geographical location is a kind of important information for satellite imagery application, but they are not fully included in ISO 19115 and 19115-2, and ISO 19130 only supports geographical positioning and sensor characteristics. In order to apply imagery data into geographical information and describe imagery metadata effectively, it is required to expand ISO metadata standards (ISO 19115, 19115-2), combining geographical location information and sensor characteristics in ISO 19130. On the basis of ISO 19115, 19115-2 and 19130, we can develop imagery metadata model and satellite data integration metadata model. By definition of common metadata sets, association and metadata expansion, abstract structure and contents can be described.

### B. Satellite metadata model

As shown in Fig .2, a complete metadata model has multiple metadata packages, including definition of main class, supporting class, expansion class and special class.

Each metadata package contains one or more metadata entities. Metadata entity consists of a series of metadata elements, which may include three components, namely core elements, special class elements and extension class elements. Here, core element defines a minimum set of metadata that each data set should contain. Elements of special class include professional elements and elements dedicated by an organization. Extension elements define extended metadata describing satellite images.

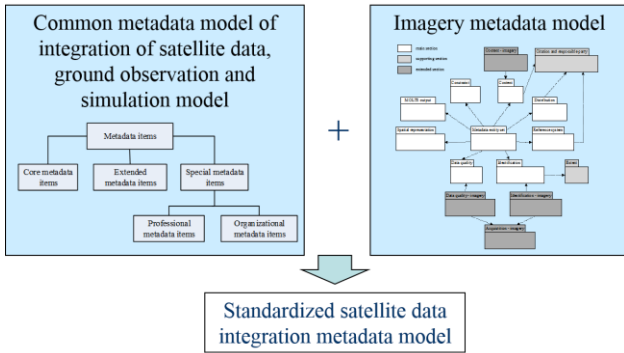


Figure 2. Metadata management of standardized satellite data integration.

#### IV. REMOTE SENSING BIG DATA ANALYSIS

By cognitive of spatio-temporal data and knowledge discovery method of massive remote sensing data, we can establish satellite remote sensing knowledge base which can reflect spatial and temporal information changes and association of ground parameters, and develop remote sensing big data analysis, including spatio-temporal queries, spatio-temporal analysis and knowledge discovery for massive satellite data.

Specifically, as shown in Fig .3, using spatio-temporal model of satellite data defined by metadata, spatial query, data analysis and knowledge discovery can be handled based on knowledge. Spatial query implements information query on attribute, geometry and spatial relationship via metadata. Data analysis implements spatial information extraction, such as automatic extraction of geographical features, image understanding, and other basic functions through metadata. The key is how to create satellite data warehouse, how to find knowledge from satellite warehouse, i.e. clustering mining, association mining, and uncertainty mining, and how to interpret and represent knowledge to user.

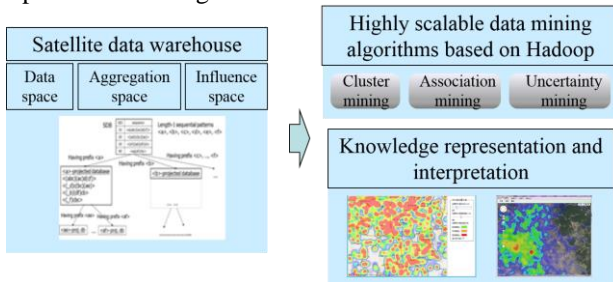


Figure 3. Remote sensing big data analysis.

##### A. Satellite data warehouse

From depth of application, data mining can be divided into three space levels as below.

- Data space. Using query functions of the existing database management system, information query based on keywords on data space is handled to achieve OLTP (On-Line Transaction Processing).
- Aggregation space. Obtain spatio-temporal object sets with spatial and temporal constraints from spatio-temporal data. Using aggregation operations (like count, sum, average, max, min etc.), and combining multi-dimensional analysis and

statistical analysis, OLAP (Online Analytical Processing) is achieved in this space in order to provide to statistical analysis for decision-making.

- Influence space. According to similarity clustering, we can discover association, similar sequence and structure pattern, and useful implicit information from data.

##### B. Cluster mining

We can search for location clusters related to non-space description, i.e. using efficient algorithm for spatio-temporal clustering analysis, and using inductive method for searching for different clustering types of data description.

$K$ -means algorithm [14] was proposed by Mac Queen in 1967. The idea of the algorithm is to find  $K$  cluster centers  $c_1, c_2, \dots, c_K$ , making square of distance minimization between each data point  $x_i$  and its nearest cluster center  $c_v$ . The inadequacy of  $K$ -means algorithm is that it is usually terminated after obtaining a local optimum value upon termination, thus it is only suitable for numerical data clustering and for those data set with convex of clustering results. To overcome the limitations of  $K$ -means algorithm, it is required to make improvements on it for clustering attribute data, making temporal information and high-dimensional data sets have good characteristics.

##### C. Association mining

Data association is a kind of important knowledge which can be discovered. Association among data can be represented by association rules in form of  $A_1 \wedge A_2 \wedge \dots \wedge A_i \rightarrow B_1 \wedge B_2 \wedge \dots \wedge B_j$ . If  $B_1, B_2, \dots, B_j$  appear, then  $A_1, A_2, \dots, A_i$  should occur, i.e. there is an association between  $A_1, A_2, \dots, A_i$  and  $B_1, B_2, \dots, B_j$ . Association analysis identifies spatio-temporal topological relationship, such as connection, adjacent, symbiosis and containing using association rules and induction techniques.

The classical algorithm of association rule mining is *Apriori* [15]. The basic idea of *Apriori* is to scan database to generate a candidate set of data items. Number of each candidate data that occurs is calculated based on pre-specified minimum support in order to generate data itemset  $\{I_1, I_2, \dots, I_N\}$ , and then produce rules from data sets. But it is required to scan database for multiple times in order to discover association rules, so  $I/O$  overhead of the algorithm is large and is not efficient. We can use extended *Apriori* method to achieve association rules. Large set of data items can be discovered from data sets. Strong rules can be generated from data items. And strong ruleset are then centralized. It is needed for the algorithm to define measurement degree for spatial association (adjacent) relationships.

##### D. Uncertainty mining

When people make judgment, forecasting or decision-making, information in problem domain are usually incomplete, inaccurate or vague.

As an intelligent data analysis tool, rough set theory are usually applied to knowledge acquisition and knowledge representation of such uncertainty. Therefore, we can handle uncertain data mining through rough set theory. Through representation of set of concepts of domain

knowledge  $A=\{a_1, a_2, \dots, a_n\}$ ,  $a_i \in A$ , object set  $O=\{o_1, o_2, \dots, o_m\}$  is formed, here,  $o_i \in O$ . Based on classification set  $C$  and object set  $O$ , a nuclear  $in(C, O)$  and an envelope  $te(C, O)$  can be generated after a certain operation on  $P$  and  $E$ , and therefore a rough set  $[in(C, O), te(C, O)]$  is established to constitute uncertainty interval, using upper limit and lower limit as a measurement of uncertainty.

#### E. Knowledge representation and interpretation

The most important application is that user is able to understand knowledge to be discovered. This requires knowledge should not be limited to conventional numerals or symbols, but rather a more easily understood manner, such as tables, histograms, scatter plots or natural language. In particular, more attention should be procedure and distribution of time and space in decision-making, which requires a real-time visual integration environment for analysis and decision. Data visualization makes interaction and control on data by way of multi-dimensional means on data representation, variable angle viewing, dynamic links and virtual reality, which can interactively analyze complex spatio-temporal relationship and dynamic trends.

#### V. DISTRIBUTED HIGH-PERFORMANCE SATELLITE DATA-CLOUD PLATFORM

Based on shared knowledge base, collaboration technology of multiple data centers can be applied to address the problem of data storage and sharing of distributed massive satellite data, so that a distributed high-performance satellite data-cloud platform can be established via cloud computing [16] that allows users access valuable data and information directly from different nodes.

Specifically, through master meta database and manipulation tools from master node of satellite data management center, users are guided to find information they need and are provided clues of detailed information. And then, through secondary meta database located in each sub-node, users can be also guided to know further information to determine their required data content, access way and means. For internal users, by meta database and manipulation tools, they can query information on other sites, and also can manage their own metadata database for maintenance. For external users, through metadata and browsing tools, they can discover information and outline or know the detailed information, and can access them through the appropriate channels. At the same time, information sharing metadata should also include detailed description of service interface of use of satellite data, establish geospatial data exchange network, and provide portal for users to enter data query and data analysis.

#### VI. CONCLUSION

Facing massive remote sensing satellite ground formation, effective method of data integration between multiple sources of geospatial information is studied. The paper focuses on the effective integration approach to integrate various heterogeneous data, like satellite data, ground observation and simulation models etc., and

integrate different data archiving systems. A technical framework is proposed for satellite data integration system to solve the issues of distributed storage and data format conversion and interoperability for massive satellite data. The proposed framework can provide an effective solution to distributed storage, data format conversion and interoperability for satellite remote sensing big data.

#### ACKNOWLEDGMENT

The paper is supported by the Open Research Fund of the Academy of Satellite Application under grant No.2014\_CXJJ-YG\_02, and the Fundamental Research Funds for the Central Universities under grant No.2042014kf0297.

#### REFERENCES

- [1] D. Li, and Q. Tong et al, "Several Cutting-Edge Scientific Problems of High-Resolution Earth Observation," *Scientia Sinica Terrae*, Vol. 42, No. 6, 2012, pp. 805-813.
  - [2] Y. Wang, "International Development of New Generation of Earth Observation Systems," *Advances in Earth Science*, Vol. 20, No. 9, 2005, pp. 980-989.
  - [3] D. Li, "Opportunities for Geomatics," *Geomatics and Information Science of Wuhan University*, Vol. 29, No. 9, 2004, pp. 753-756.
  - [4] G. Li, and Y. Wu, "Research of Earth Observation Science in Big Data Era," *China Computer Society Newsletter*, Vol. 9, No. 9, 2013, pp. 27-31.
  - [5] D. Sui, M. Goodchild, and S. Elwood, "Volunteered Geographic Information, the Exaflood, and the Growing Digital Divide," *Crowdsourcing Geographic Knowledge*, 2013, pp. 1-12.
  - [6] ISO, "ISO/TC 211 - Geographic Information/Geomatics," available at [http://www.iso.org/iso/home/store/catalogue\\_tc/catalogue\\_tc\\_br\\_owse.htm?commid=54904](http://www.iso.org/iso/home/store/catalogue_tc/catalogue_tc_br_owse.htm?commid=54904).
  - [7] OGC, "The OpenGIS Abstract Specification," available at <http://www.opengeospatial.org/standard/as,2005>.
  - [8] China High-Resolution Earth Observation System, available at <http://www.sastind.gov.cn/n25770/index.html>.
  - [9] M. Craglia, K. Bie, and M. Pesaresi et al., "Digital Earth 2020: Towards the Vision for the Next Decade," *International Journal Digital Earth*, Vol. 5, 2012, pp. 4-21.
  - [10] X. Gu, "Development of China's Industry of Satellite Application," Academic Report, Wuhan University, 2010.
  - [11] X. Gu, "Status and Prospect of Development of China Space Observation System," Academic Report, Wuhan University, 2010.
  - [12] R. Xie, and R. Shibasaki, "The establishment of satellite data integration system," *GIM International*, Vol. 20, No. 12, 2006, pp. 21-23.
  - [13] R. Xie, and R. Shibasaki, "Metadata Development for the Integration of CEOP Satellite-Observation Data," *Journal of Meteorological Society of Japan*, Vol. 85A, 2007, pp.487-517.
  - [14] C. Ordonez, and E. Omiecinski, "Efficient Disk-based K-means Clustering for Relational Databases," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 16, No. 8, 2004, pp. 909-921.
  - [15] R. Agrawal, and R. Srikant, "Fast Algorithms for Mining Association Rules in Large Databases," *Proceedings of the 20th International Conference on Very Large Data Bases*, Santiago, Chile, September, 1994, pp. 487-499.
- T. Erl, R. Puttini, and Z. Mahmood, "Cloud Computing: Concepts, Technology & Architecture," Prentice Hall, ServiceTech Press, 2013.