UNISAT. The Technology for Development of United Systems of Maintaining Extra Large Distributed Archives of Heterogeneous Satellite Data

Andrey A. Proshin, Evgeny A. Loupian, Alexandr V. Kashnitskii, Ivan V. Balashov

Space Research Institute RAS, Moscow

Abstract. Rapid development of satellite Earth remote sensing led to a significant increase in requirements for systems for maintaining satellite data archives. The article describes the UNISAT technology, designed to build systems for maintaining extra large distributed archives of heterogeneous satellite data, providing dynamic generation of data at the user's request and a wide range of tools for remote analysis and data processing.

Keywords: UNISAT Technology

1   Introduction

Rapid development of satellite remote sensing systems in recent decades has led to an explosive increase in the volume of satellite data obtained from a lot of various observation instruments [1]. The field of application of earth remote sensing data is expanding, as the data are now widely used for a variety of research and applied tasks in the natural environment and anthropogenic objects is expanding area. In turn, all this leads to a significant increase in the requirements for systems that provide satellite data processing, and in particular, to systems for maintaining satellite data archives which provide the back end for the data analysis. One of the main requirements for modern satellite data archiving systems is to support the operation of heterogeneous satellite data obtained by observation devices with different technical characteristics (observation frequency, spatial resolution, repeatability of observations, etc.). This leads to the need for unification of data archiving procedures, development of a common database structure, and implementation of common software interfaces to access the variety of satellite data types. Another important requirement is the support of extra large distributed satellite data archives, which enables the joint operation of the archives located in a number of satellite data acquisition and storage centers making a single information resource. Thus, the users access the data, no matter where they are physically located.

The requirements for data access services have also changed dramatically in recent years. Previously, the users of satellite data were mostly satisfied with obtaining raw data for use in their processing and analysis systems, but now they are increasingly interested in the possibilities of accessing the ready-made data products at a number of processing levels [2], moreover, the number of such data products required to solve specific problems is constantly growing. It should be noted that, since satellite data tend to take a lot of disk space, the storage of all possible data products derived from the same source data becomes impractical and, in many cases, technically impossible. The reasonable way out of this situation is to provide users with access to "virtual" data products, i.e. products that are dynamically built from source data in real time. The key advantage of this approach is the ability to expand the list of data products available to users without the need for mass processing of the data archives.

In our opinion, one of the most urgent development directions of modern satellite data access systems is the implementation of various tools for satellite data processing and analysis, which were previously available only in specialized desktop applications. The data analysis tools implemented through these interfaces enable the processing of large amounts of available satellite information, using the capacity of the data centres. The most prominent representatives of such systems, in our opinion, are Google Earth Engine [3] (https://earthengine. google.org) and the "Vega-Science" system [4] implemented in framework of the "IKI-Monitoring" center for collective use. [5].

This article gives a brief description of the key features of the UNISAT technology developed in IKI RAS [6], designed to create unified systems for maintaining extra large distributed archives of heterogeneous satellite data. In turn, it is based on the use of technologies and software that have been developed in IKI RAS in recent decades to build a wide range of various archives of satellite data [7,8]. UNISAT technology implements the common approach to processing a variety of satellite data types, which differ both in spatial resolution and the scheme of data storage. An important advantage of the developed technology is the support of the "virtual data products" mechanism, i.e. products that are dynamically generated upon the user's request based on the processing of satellite data available in the archives. The development of this technology was also focused on supporting a wide range of various tools for remote analysis and processing of satellite data.

2 General Architecture

Years of experience in creation and operation of various systems for satellite data access enabled us to develop a common architecture for building systems for maintaining satellite data archives. Its key elements make up the new unified structure of the database for storing heterogeneous satellite data, as well as the structure of the reference database containing detailed information about satellites, observation instruments, implemented products, as well as the schema for building "virtual" data products.

The General architecture of the node of the distributed archive management system UNISAT is shown in Fig.1. Data can be imported to the archives both the satellite data processing subsystem and from external suppliers of satellite data. The left part of the diagram represents the software components responsible for satellite data archiving and data exchange with other information centers. The middle part of the diagram contains the reference database unisat_catalog, the unisat database, containing metadata available in the archive data, and the connected file storage, which physically contains the satellite data files. The right part of the diagram shows the main services designed to provide access to the data in the archive. The cartographic WEB interface is the main tool for providing users with access to both the satellite data itself and the services designed to analyze them. In the diagram, the dotted arrows show the requests for data or metadata, solid arrows stand for the metadata flow, and hollow arrows mean the data flow. External components that are not being a part of the archiving system are shown in dotted shapes.
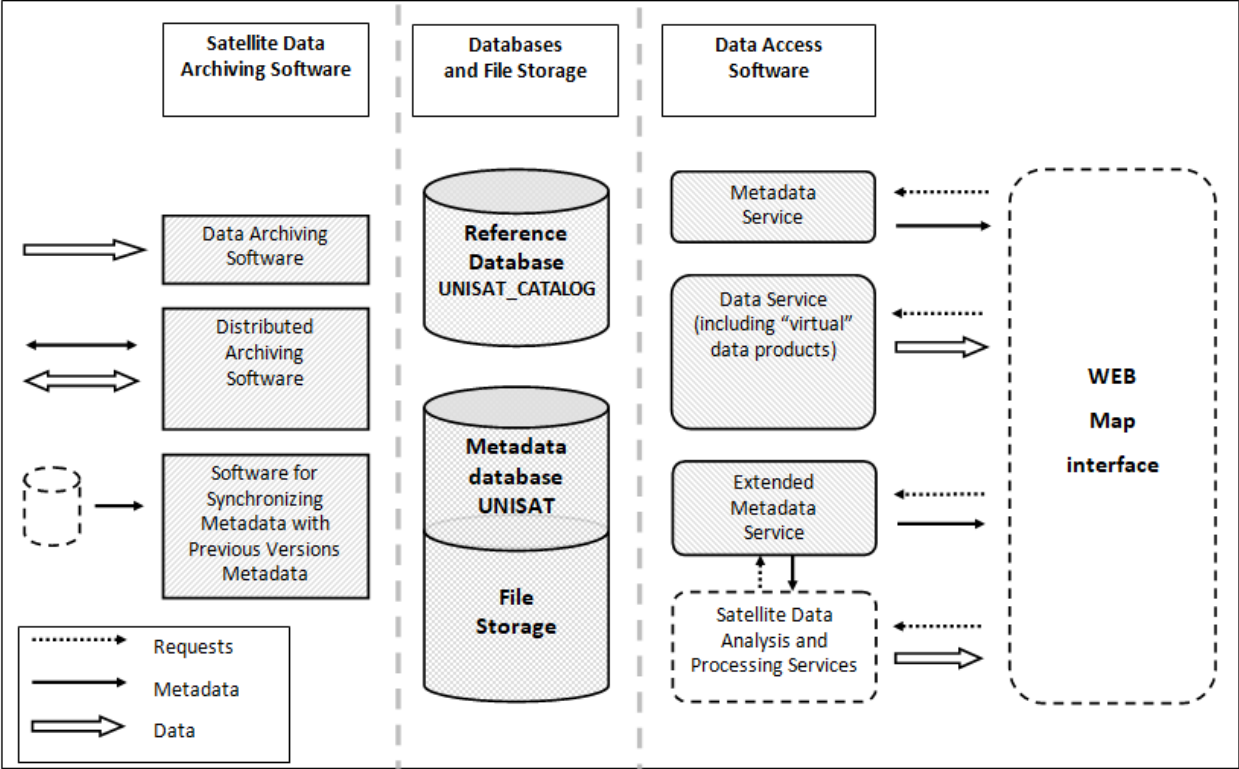

Figure. 1. The General architecture of the node of the distributed archive.

Fig. 2 represents the schematic design diagram of the satellite data distributed archive implemented with use of the presented technology. The diagram shows the main data and metadata flows implemented within the distributed archive. Data can be fed into archives from both external data centers and local reception stations. For each information centre, a policy for exporting data or metadata to the remaining centers of the distributed archive can be defined individually, but in the simplest version, each centre contains all the information about the data available in the distributed archive. Each center can implemented its own subset of satellite data archives, but the reference database unisat_catalog always remains synchronized with the central server. The main advantage of the presented implementation of the distributed satellite data archive is a high degree of flexibility in determining what types of metadata and data should be transferred between the data centers in framework of the distributed archive, which is practically impossible with the use of standard database replication tools.
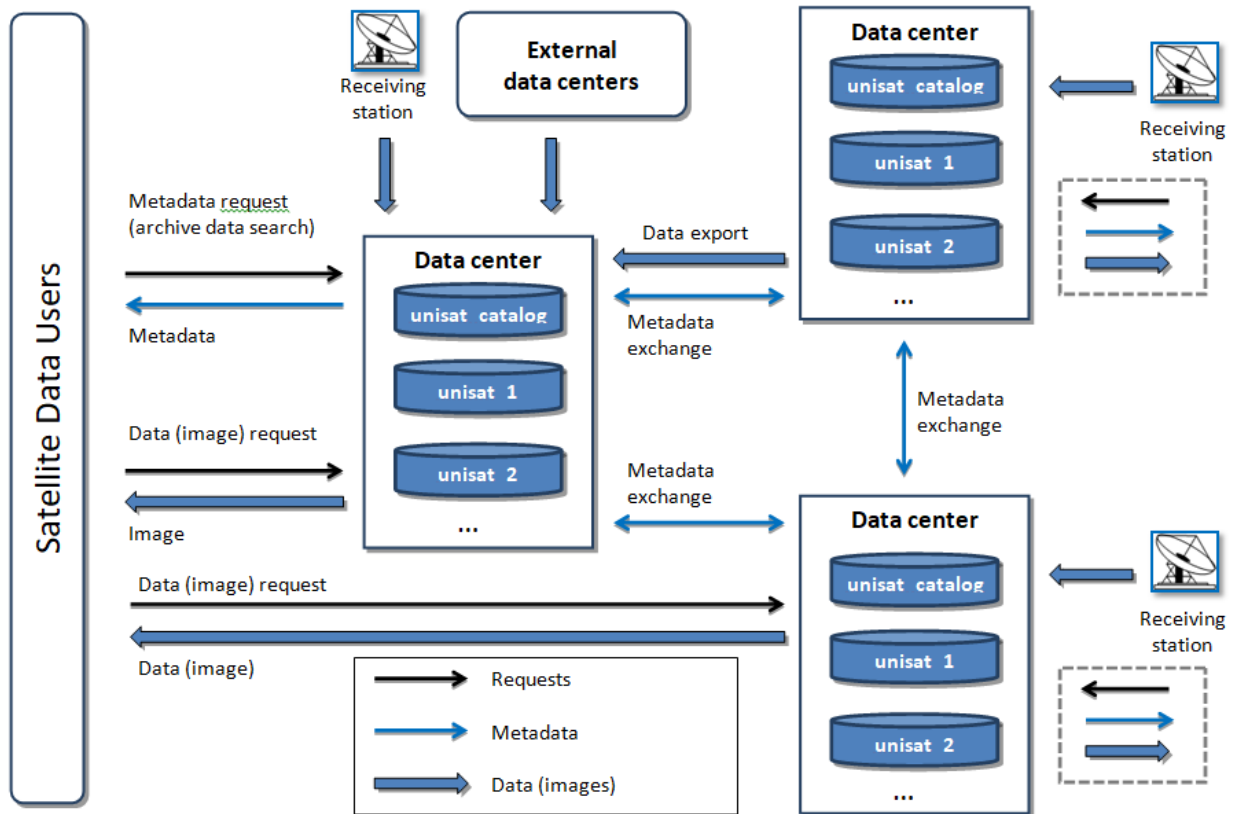

Figure. 2. Schematic design diagram of the distributed data archives.

3   The Unified UNISAT Database Structure for Satellite Data Storage

To define the unified database structure, let's introduce a number of terms. A session is a set of data uniquely identified by the fields: date and time (**dt**), satellite (**satellite**), station (**station**), instrument (**device**). The fragment is the spatial part of the session. The session can consist of a single fragment or a set of fragments. The UNISAT database has two main tables for storing metadata: the table **fragments** intended to describe the fragments, and the table **fragment_products** for definition of the data products specific to the respective fragments (see Fig. 3). In the figure, the primary keys of the tables are shown in bold, and the fields of the unique key are shown in italics.

In case a session consists of one fragment, one record is added into the **fragments** table indicating the type of fragment "**single_fragment**". And in case the session consists of many fragments, each of them corresponds to a separate record with the fragment type "**fragment**" and a unique fragment number, but it also requires that fragments of the type "**products_contour**" are automatically generated to define the outer contours of products or product groups in framework of the session. The **scale_level** field in the **fragment_products** table is used to identify coarse versions of data products used to accelerate the display of images at scales exceeding the resolutions of the satellite data itself. It is significant that the same products can have different spatial partitioning at different scales.
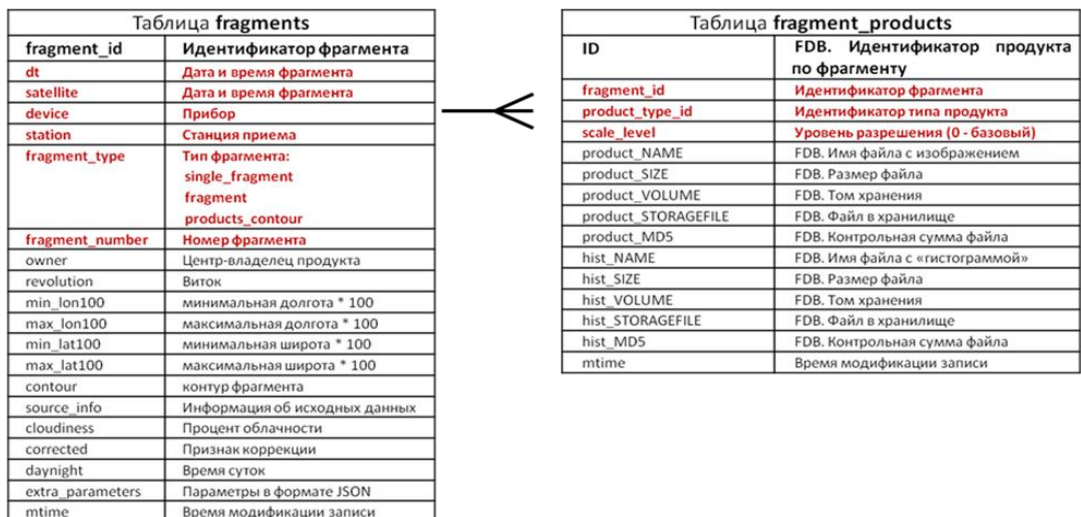
Таблица fragments

fragment_id	Идентификатор фрагмента
dt	Дата и время фрагмента
satellite	Дата и время фрагмента
device	Прибор
station	Станция приема
fragment_type	Тип фрагмента: single_fragment fragment products_contour
fragment_number	Номер фрагмента
owner	Центр-владелец продукта
revolution	Виток
min_lon100	минимальная долгота * 100
max_lon100	максимальная долгота * 100
min_lat100	минимальная широта * 100
max_lat100	максимальная широта * 100
contour	контур фрагмента
source_info	Информация об исходных данных
cloudiness	Процент облачности
corrected	Признак коррекции
daynight	Время суток
extra_parameters	Параметры в формате JSON
mtime	Время модификации записи

Таблица fragment_products

ID	FDB. Идентификатор продукта по фрагменту
fragment_id	Идентификатор фрагмента
product_type_id	Идентификатор типа продукта
scale_level	Уровень разрешения (0 - базовый)
product_NAME	FDB. Имя файла с изображением
product_SIZE	FDB. Размер файла
product_VOLUME	FDB. Том хранения
product_STORAGEFILE	FDB. Файл в хранилище
product_MD5	FDB. Контрольная сумма файла
hist_NAME	FDB. Имя файла с «гистограммой»
hist_SIZE	FDB. Размер файла
hist_VOLUME	FDB. Том хранения
hist_STORAGEFILE	FDB. Файл в хранилище
hist_MD5	FDB. Контрольная сумма файла
mtime	Время модификации записи

Figure. 3. The structure of the main UNISAT database tables.

The developed structure made it possible to process efficiently the data both in form of separate fragments and in form of integrated products consisting of many thousands of separate fragments. The advantages of the described structure also include flexible support for additional spatial partitions at coarse scales, which allows to accelerate the formation of the required images at a given scale.

4 The Structure of the unisat_catalog Common Reference Database

The database is designed to maintain all the necessary reference information about receiving stations, satellites, characteristics of satellite instruments, etc., and also contains information about the types of products stored in the archives and the schema for obtaining "virtual" information products on their basis. Table 1 below represents the composition and purpose of the main tables of the **unisat_catalog** database, indicating the type of reference information stored in them. The key difference of the proposed structure of the "reference" database is the integration of common information on satellites, instruments and their respective data types with the information necessary to obtain "virtual" data products. At the same time, the implemented services for obtaining extended metadata allow to provide information necessary for the implementation of tools for remote analysis and processing of satellite data.

Table 1. The composition and purpose of the unisat_catalog database tables

Information type	Table	Purpose
General reference information	satellite	information on satellites
	satellite_device	information on instruments installed on satellites
	device	information on satellite instruments
	band	information on satellite instrument channels
	station	information on satellite data receiving stations
	center	information on data centres
Description of the types of products stored in the archives	product	description of types of data products
	product_cases	information on the types of products built from the data of specific satellite instruments
	product_level	information on "coarse" product scales
	channel	information on channels of information products
Schema for obtaining "virtual" products based on the processing of products in the archive	vproduct	description of virtual product types
	vproduct_cases	variants for the implementation of virtual products depending on the types of satellite instruments and available data in the archive
	vchannel	schema for obtaining channels of the virtual products

5   Conclusion

Currently, the UNISAT technology is being successfully used in the implementation of a number of information systems for satellite data access developed by IKI RAS in cooperation with other organizations. Among them, it is particularly worth noting the "IKI-Monitoring" satellite data center for collective use [5], which provides direct access to satellite data for more than 30 different Earth observation instruments, with the total amount of available data now significantly exceeding 2 petabytes. The same technology is used to maintain archives of the unified satellite data processing system of "SIC "Planeta" [9], which contain data of both Russian and foreign satellite systems of Earth remote sensing. In the future, it is planned to provide effective support for a wider range of tools for remote analysis and processing of satellite data.

References

[1] Loupian E.A., Bourtsev M.A., Proshin A.A., Kobets D.A. Evolution of remote monitoring information systems development concepts // Actual Problems of Remote Sensing of the Earth from Space. 2018. Vol. 15. No. 3. P. 53-66. DOI: DOI: 10.21046/2070-7401-2018-15-3-53-66.

[2] Loupian E.A., Savorsky V.P. Basic products of Earth Remote Sensing Data Processing // Actual Problems of Remote Sensing of the Earth from Space. 2012. V. 9. № 2. P. 87-97.

[3] Moore R. T., Hansen M. C. Google Earth Engine: a new cloud-computing platform for global-scale earth observation data and analysis // AGU Fall Meeting Abstracts, 2011. Vol. 1. P. 2.

[4] Tolpin V.A., Loupian E.A., Bartalev S.A., Plotnikov D.E., Matveev A.M. Possibilities of agricultural vegetation condition analysis with the "VEGA" satellite service // Atmospheric and Oceanic Optics. 2014. Vol. 27. No. 7 (306). P. 581-586

[5] Loupian E.A., Proshin A.A., Bourtsev M.A., Balashov I.V., Bartalev S.A., Efremov V. Yu., Kashnitskiy A.V., Mazurov A.A., Matveev A.M., Sudneva O.A., Sychugov I.G., Tolpin V.A., Uvarov I.A. IKI center for collective use of satellite data archiving, processing and analysis systems aimed at solving the problems of environmental study and monitoring // Actual Problems of Remote Sensing of the Earth from Space. 2015. Vol.12. No 5. P. 263-284.

[6] Proshin A.A., Loupian E.A., Balashov I.V., Kashnitskiy A.V., Bourtsev M.A. Unified satellite data archive management platform for remote monitoring systems development // Actual Problems of Remote Sensing of the Earth from Space. 2016. Vol. 13. No. 3. P. 9-27. DOI: 10.21046/2070-7401-2016-13-3-9-27.

[7] Loupian E.A., Balashov I.V., Bourtsev M.A., Efremov V. Yu., Kashnitskiy A.V., Kobets D.A., Krasheninnikova Yu. S., Mazurov A.A., Nazipov R.R., Proshin A.A., Sychugov I.G., Tolpin V.A., Uvarov I.A., Flitman E.V. Development of information systems design technologies // Actual Problems of Remote Sensing of the Earth from Space. 2015. Vol.12. No 5. P. 53-75.

[8] Efremov V. Yu., Loupian E.A., Mazurov A.A., Proshin A.A., Flitman E.V. A`Technology for Construction of Automated Systems for Satellite Data Storage // "Actual Problems of Remote Sensing of the Earth from Space: Physics, Methods and Technologies for Monitoring of Environment and Hazardous Phenomena and Objects", Moscow, Polygraph-Service, 2004. P. 437-443.

[9] Loupian E.A., Milexin O.E., Antonov V.N., Kramareva L.S., Bourtsev M.A., Balashov I.V., Tolpin V.A., Solovyev V.I. System of operation of joint information resources based on satellite data in the Planeta Research Centers for Space Hydrometeorology // Russian Meteorology and Hydrology. 2014. Vol. 39. Issue 12. P. 847-853. DOI: 10.3103/S1068373914120103.