Information and computing maintenance for the regional satellite operational monitoring system

O E Yakubailik¹, E A Pavlichenko² and V Y Romas'ko³

¹ Institute of Computational Modelling SB RAS, Krasnoyarsk, Russia
² Federal Scientific Center «Krasnoyarsk Scientific Center of the Siberian Branch of the Russian Academy of Sciences», Krasnoyarsk, Russia

E-mail: oleg@icm.krasn.ru

Abstract. The article discusses the experience of supporting a regional operational satellite monitoring system. Based on NASA IOPP free & open source software, a set of programs has been created to automate the reception and automated processing of incoming satellite information. The development is primarily focused on data coming from the satellite receiving station, however, it also deals with the interpretation of data received via the Internet. A three-stage scheme of distributed processing and data conversion is implemented. The output goes to the local satellite data storage. A collection of specialized multi-channel raster images is also being formed for quick visualization of images in a geographic information web system. The current version of the system performs processing from eight different sensors installed on several satellites.

1. Introduction
One of the steps to address the urgent tasks of ensuring the effective use of the results of space activities in the Krasnoyarsk Territory was the acquisition and commissioning in 2017 of a new satellite receiving complex of the FRC KSC SB RAS based on the UniScan-36 SCANEX station [1], which provides the ability to receive remote sensing data from all relevant low-orbit meteorological satellites of medium spatial resolution (American TERRA, AQUA, Suomi NPP, NOAA-20, Chinese FengYun-3). Despite the possibility of obtaining various remote sensing information from different sources on the Internet, an autonomous station has advantages in efficiency, independence from various interstate sanctions, and flexibility in automating applications for regional needs, and thus remains attractive as a source of remote sensing data at the regional level. Based on such equipment, in recent years there has been a trend towards creating problem-oriented regional operational satellite monitoring systems for solving various thematic problems in the field of ecology, agriculture, fire hazard forecasting, and many other areas of activity [2, 3].

The main problem of the functioning of autonomous receiving stations and monitoring systems built on their basis is the automation of the processing, archiving and cataloging of data. These data are generated during the reception by the station, and they can also come from third-party sources of information via the Internet. Typically, high-resolution satellite images or data from some specific satellites are processed by information providers and then transmitted over the Internet. Managing the processing of a large flow of information is quickly becoming the main problem of the station, since the daily amount of data reaches large volumes. Despite the fact that the delivery of the satellite complex
includes a basic set of software for processing received satellite data, the practice and accumulated experience in operating the station have shown the feasibility and effectiveness of using additional software for the tasks of operational satellite monitoring.

The aim of the work was to ensure maximum automation of the processing of archived and operational satellite data received from Terra / Aqua and Suomi-NPP / NOAA-20 meteorological spacecraft using the standard (unchanged) IPOPP software package, as well as the creation of output (thematic) data archives for further research.

2. Materials and methods

IPOPP (International Planetary Observation Processing Package) is a free and open source software package for processing Earth remote sensing data from several NASA (National Aeronautics and Space Administration) meteorological satellites of the United States. IPOPP is supported by the Direct Readout Laboratory (DRL), which links remote sensing programs (such as TERRA, AQUA, SNPP and JPSS-1 and 2) and users of autonomous receiving stations that are not directly involved in these programs. The website of the DRL laboratory has become a platform for information interaction between software developers and the user community of receiving stations and serves to exchange information and technologies [4].

The IPOPP package is distributed as a ready to use software product that is installed in the environment of the Centos 7.0 Linux operating system. It combines various Science Processing Algorithms (SPA), such as MOD14 – a package for detecting forest fires according to MODIS, VFIRE375 – a package for detecting fires according to VIIRS with a resolution of 375 m, CLOUDMASK – a package for computing a cloud mask, AEROSOL – a package for calculating various aerosol properties, and other packages [5, 6]. Because most algorithms are interconnected and use the results of other algorithms, all packages are combined into one framework to simplify processing and visualize the data processing process. The processing process is an inverted tree structure, processing starts from the input point, which is the processing of raw data, and gradually branches and parallelizes. Data processing using SPA algorithms is automated, but operator intervention is required to start the data processing process and obtain ready-made thematic products. To process large amounts of archival data (for example, data for 3 years taking into account the reception of 8 scenes per day from several satellites are presented in more than 8000 raw data files) or to process continuously incoming operational data, complete automation of the processing process is required, which should exclude operator intervention.

The automation process is divided into three parts. The first part is the process of waiting for new files in the raw data directory and adding a list of new data to the processing queue. The second part consists of processing the data of the processing queue and deleting the processed data from it. And the last one is copying the finished results from the output catalog of finished IPOPP products to the archive. Before starting automatic data processing, you must manually select the list of SPA data processing algorithms and configure the list of ready-made thematic products extracted from the output directory to the archive [7].

Automation of the processing of satellite data is performed in the bash shell of the Centos Linux operating system. All automation programs are aimed at minimal changes to the original IPOPP package, so that when updating individual SPA packages or the entire IPOPP software package, re-configuration of packages or changes in the automation programs themselves is not required. To monitor the raw data catalog, the inotify-tools software package is used, which allows you to track file system changes and receive a list of only new or changed files. Online raw data is added to the top of the processing queue. The processing of the raw data queue is in order, due to which the priority processing of operational data is achieved. Archived data is added to the end of the processing queue and processed at the end of the operational data queue. Processing data from the queue and copying the finished data to the archive occurs according to the planned schedule in the cron program. A data archive is created with a given storage structure, which facilitates the search and access to data. Processed operational data
is deleted using IPOPP tools, with the interval specified in the package settings, usually 14 days. This interval is required for multi-day SPA algorithms.

Using an automation package in the Krasnoyarsk Regional Center, real-time data is processed, and batch processing of portions of archive data is also possible. An example of visualization of the results of SPA algorithms is presented in figure 1.

![Figure 1](image.png)

**Figure 1.** The results of various IPOPP algorithms: NDVI and EVI (on the left), Fire Thermal Points 375 m (in the center) against the Earth's surface in natural colors, Aerosol Index map (on the right).

In the process, identified shortcomings that require further refinement. The IPOPP software requires additional ancillary data that is downloaded from the NASA server directly during batch processing. Such data loading significantly slows down the processing, since the size of one additional data packet is approximately 2 Gb. In the case of reprocessing of archived data, the same data is reloaded. To speed up processing and avoid re-downloads, changes were made to the ancillary data download program. Now the data is downloaded at night, the downloaded data is not deleted, and when reprocessing the archive data, it is reused.

Another important task of providing an information-computer system for the reception, storage, processing and visualization of remote sensing data is the cataloging of satellite data. The continuous operation of the station quickly creates a huge number of different pieces of data. And if the time search can be somehow organized by grouping data into catalogs by day, month, year, etc., then searching for information on the spatial coverage and cloudiness of the images immediately becomes a problem. To solve it, it is necessary to organize and maintain a spatial database of all available archived images, including, in addition to the metadata of the images, overview images. Overview images are reduced copies of the source images, composed of the most informative satellite imagery channels, specially designed to search for monitoring targets and assess the quality of their visibility in this image [8, 9].

The procedure for preparing overview images and collecting metadata can be fully automated, since the composition and format of the source data does not change over time. Despite the obvious need for the program for this, fully finished software for these purposes is absent both in the supply of the station and in the software market. To solve the problem of cataloging archive information, a program was developed for organizing a fully automatic process for preparing multichannel overview images in GeoTIFF format and associated metadata based on remote sensing data taken online by the satellite receiving complex of the Federal Research Center of the Siberian Branch of the Russian Academy of Sciences obtained by ordering surveys of target areas in Roskosmos or downloadable from public resources. The target consumer of these multi-channel overview images is a satellite data cataloging system. A side goal is the organization of automatic control of archiving of operational data received at the Uniskan-36 station, the stream of which reaches 90 Gb per day.

Full automation of data processing is implemented in the form of organizing a program as a service on a workstation running Windows 10. Unlike regular programs, services are constantly running and run when the OS starts, which allows you to implement any activity algorithm: periodic, event, or any other. The program messages are output to the Windows own log using the standard API for these purposes. The configuration of the service is completely collected in one text file, which is read during
each cycle of service activity, which allows you to change almost the entire configuration on the fly. To monitor the operation of the service, it includes an embedded HTTP server that implements a web interface for monitoring the execution of tasks by the service, which includes the following: 30 recent messages of selected importance, a list of tasks in the "portfolio", current tasks to be performed, free space for results and temporary files.

Service is a multi-threaded program organized in the paradigm of the so-called “portfolio of tasks” [10]. This paradigm implies the presence of several flows: “workers”, continuously and simultaneously performing the “tasks” that they take from the common “portfolio”. Tasks are put into the “portfolio” by the employees themselves in the process of performing other tasks, as well as specific task flows. The service has one task factory that periodically scans the source directories and generates data processing tasks, and several (the number is specified in the configuration) workflows.

The service configuration is 3-level and consists of the following components: (1) general parameters, (2) groups of parameters for processing specific satellites, (3) a list of instruments (sensors) installed on the satellites with their own composition of channels and other settings. Such an organization is determined by the nature of the source data, which vary greatly from one another depending on the instrument. For some parameters, “inheritance” of values from the upper levels is used, for example, for the projection of overview images, which in fact is the same for all satellite data. The service also widely uses templates for specifying file names: regular expressions for parsing and unambiguous identification of source file names and macro substitution in result file name templates. All this tools together allows you to build a flexible file management system. For a fair data processing strategy for all configured groups, a single cycle of viewing and generating processing tasks is applied, which, when executed, scans the source directories in search of new files sequentially for all groups.

3. Results and discussion
At present, 8 radiometer data processing algorithms have been implemented: (1) MODIS of the Terra and Aqua satellites, (2) VIIRS from Suomi NPP and NOAA-20 satellites, (3) MSI from Sentinel-2A and -2B satellites, (4) OLI and TIRS sensors from Landsat-8 satellite, (5) MSS and (6) PSS sensors from Kanopus-V, BKA, Kanopus-V-IK, Kanopus-3-6 satellites, (7) Geoton Panchrome and (8) Multichannel data from Resurs-P No. 1-3 satellites. Factorization of all 8 algorithms revealed the following typical steps for completing a task:

1. Download to memory or the formation in memory of images from the source data.
   An image is formed in memory either by loading the necessary channels (the most common option), or by collecting from separate fragments (Geoton), or by extracting data from multidimensional arrays of the HDF format (MODIS and VIIRS). Data of a very large volume (Geoton) at the same step is reduced to the minimum specified pixel size.

2. Calibrate or retrieve calibration information.
   If necessary, data is calibrated (MSI, OLI and TIRS). Some devices do not have calibration at all (MSS and MSS) or do not always have it (Geoton)

3. Correction to the height of the Sun (if possible and necessary).
   If necessary, the data is adjusted to the height of the Sun (MODIS, OLI). This correction is necessary if individual images form a continuous coating, that is, they must fit together without gaps or overlap each other.

4. Recalibration for subsequent output to GeoTIFF.
   The image is converted to a standard representation of pixels, directly suitable for recording in GeoTIFF. For example, the albedo is represented as a 16-bit unsigned integer at 0.01%, and the radiation temperature at 0.01 K.

5. Collection of metadata – shared and specific for each channel.
   Metadata is extracted from the source data, including the image boundaries in projective coordinates and the image outline in geographical coordinates in the OGC WKT view.
6. **Converting an image to desired map projection.**
   The image is converted to the desired map projection. Conversion to the projection is done according to two different algorithms, depending on whether the original image has a scan structure with a butterfly effect (MODIS and VIIRS) or not (all other devices).

7. **Cyclic output GeoTIFF-s with a decrease in size by 2 times on each cycle.**
   A pyramid of overview images is created in a cycle with a doubling of the pixel size on each cycle until a predetermined limit is reached. Data output is done with a transaction simulation - either all files are written in full, or none of them are written - to exclude incomplete output due to lack of disk space.

8. **Distribution of source files into archive directories (if required).**
   The operational source data received at the Uniskan-36 station are distributed in archive catalogs. High-resolution data received from Roskosmos (MSS, MSS, Geoton) or downloaded from public sources (MSI, OLI and TIRS) are processed from the archive location and do not require distribution, and to prevent their reprocessing, corresponding log files of already processed data are kept. Distributed files are checked before processing for errors detected by specific log files. The detection of such files leads to a warning and prevents processing to avoid errors that the service cannot bypass without the help of an operator.

Figure 2 shows the visualization of data from the Suomi NPP satellite, implemented through the pyramids of overview images in the satellite data catalog. The figure shows the radiation temperature in the color palette according to the calibrated data in the overview image. Along with this view, there are 4 more options, including NDVI, presented in the legend of the source data. All these representations are displayed by a combination of different spectral channels of the same overview image, which is most suitable for the current scale of the size map from the scale pyramid.

![Figure 2](image-url)

**Figure 2.** An example of using an overview image to display radiation temperature in a color scale in the satellite imagery catalog of Krasnoyarsk Regional Remote Sensing Center.

4. **Conclusions**
   The developed software makes it possible to almost completely automate the preparation of overview images and metadata for input into a spatial database that implements the search functions for images by time and space, spectral characteristics, and also according to the conditions of visibility of targets.
taking into account cloudiness. The results of these software tools are the first level of abstraction from the specific features of various remote sensing systems, which allows you to use one database for all remote sensing instruments and devices and common search principles for them.

References
[1] SCANEX UniScan™ Ground Receiving Station Online: http://www.scanex.ru/en/station/
[2] Bordogna G, Kliment T, Frigerio L, Brivio P, Crema A, Stroppiana D, Boschetti M and Sterlacchini S 2016 *ISPRS International Journal of Geo-Information* 5(5) 73
[3] Clerici M, Combal B, Pekel J, Dubois G, van't Klooster J, Skøien J and Bartholomé E 2013 *Ecological Informatics* 18 162-70
[4] Coronado P, Overton J and Brentzel K 2009 International Polar Orbiting Processing Package (IPOPP) Framework for Earth Remote Sensing Science Data Processing Online: https://directreadout.sci.gsfc.nasa.gov/links/rsd_eosdb/PDF/IPOPP_Framework.pdf
[5] Csiszar I, Schroeder W, Giglio L, Ellicott E, Vadrevu2 K, Justice C and Wind B 2013 *Journal of Geophysical Research: Atmospheres* 119 803-16
[6] Schroeder W, Olivia P, Giglio L and Csizsar I 2014 *Remote Sensing of Environment* 143 85-96
[7] Yakubailik O, Romas’ko V and Pavlichenko E 2019 *E3S Web of Conferences* 75 03003
[8] Yakubailik O, Kadochnikov A 2019 *IOP Conference Series: MSE* 516 012030
[9] Yakubailik O, Kadochnikov A, Tokarev A 2019 *CEUR Workshop Proceedings* 2527 47-51
[10] Carriero N, Gelernter D and Leichter J 1986 *Thirteenth ACM Symposium on Principles of Programming Languages Conf.* (St. Petersburg, Florida, ACM) pp 236-42