Journal of Computer Engineering
ISSN 10003428

Original Research Article
Vol 15, Issue 3, 2026 | ISSN: 1000-3428

The JRCEarth Observation Data and Processing System

E. Kastrati, M. Vasiljevic, A. M. C. Sanches, L. M. De Souza, P. A. C. Ferreira, C. R. L. Pimenta, D. R. Silva, E. T. L. Marques, M. V. Nogueira

University of Amsterdam, Department of Computer Science, Science Park 904, 1098 XH Amsterdam, Netherlands; Federal University of Rio de Janeiro, Department of Informatics, Av. Brig. Faria Lima, 290, 21941-909 Rio de Janeiro, RJ, Brazil

Abstract

The JRC Earth Observation Data and Processing Platform (JEODPP) is a versatile petabyte-scale platform that serves the needs of a wide variety of projects. This is achieved by providing a cluster environment for batch processing, a webbased remote desktop access with a variety of software suites, and a web-based interactive visualisation and analysis ecosystem. These three layers are complementary and are all relying on a common hardware layer where the data is co-located with the processing services. The versatility of the platform is illustrated by a series of applications running on the JEODPP.

Indexed keywords: EOS, Docker, Jupyter, HTCondor, batch processing, interactive visualisation, deferred processing

Article History: Received: 05 December 2025 | Accepted: 18 February 2026 | Published: 09 March 2026

1. INTRODUCTION

Earth Observation is truly undergoing a big data shift following the free, full, and open availability of the data generated by the EU Copernicus programme and other initiatives. This shift motivated the development of the concept of the JRC Earth Observation Data and Processing Platform (JEODPP) to fulfill the needs of the JRC projects relying on geospatial data analysis in the context of their policy support activities [17]. The JEODPP needs to address the needs of users with very different requirements originating from domain experts that developed methods and algorithms in a variety of programming languages over the years to occasional users with little or no knowledge of programming. The JEODPP is a versatile platform that meets these requirements by following a multi-layer architecture where each layer serves the needs of specific user groups [16].

This paper summarises the main components of the JEODPP. Section 2 presents an overview of the platform architecture and details its hardware layer as well as the chosen distributed file system. The different types of data stored on the JEODPP are presented in Sec. 3. The main software layers are detailed in Sec. 4. Before concluding, an application gallery is presented in Sec. 5.

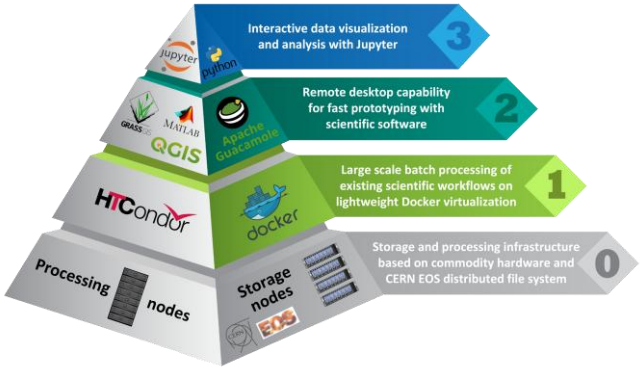

Fig. 1: The JEODPP architecture: conceptual representation in the form of a 4-layer pyramid.

2. ARCHITECTURE

A conceptual representation of the JEODPP architecture is sketched in Fig. 1 in the form of a four layer pyramid: the base layer (0) serves as a basis for the batch processing (1), the remote desktop (2), and interactive computational layers
(3)

The base layer consists of scalable commodity hardware for processing and storage servers with directly attached storage (Just a Bunch of Disks or JBODs). Currently it is equipped with 16 storage servers for a gross capacity of 1.8 petabyte and 37 processing servers for a total of 992 core CPUs. The I/O bottleneck typically observed with network attached storage is avoided by considering appropriate high speed inter-communication topology. A key component behind the storage system is the underlying distributed file system that enables each processing server to have a unified view of all the files stored in the various disks attached to the storage servers. The JEODPP relies on the EOS file system [1] developed by CERN to achieve this goal. EOS is mainly focused on low latency, high

How to cite: E. Kastrati, M. Vasiljevic, A. M. C. Sanches, L. M. De Souza, P. A. C. Ferreira, C. R. L. Pimenta, D. R. Silva, E. T. L. Marques, M. V. Nogueira (2026). The JRCEarth Observation Data and Processing System. Journal of Computer Engineering, 15(3), 176–180. DOI: https://doi.org/10.5281/zenodo.19360745

availability, ease of operation and low total cost of ownership. It is in production at CERN since 2011 and is managing more than 140 petabytes of raw disk space as of 2015 [12]. It allows for a flexible management of replica (redundancy) levels and works well in mixed hardware configurations. Files are accessed via the Institutional Platforms

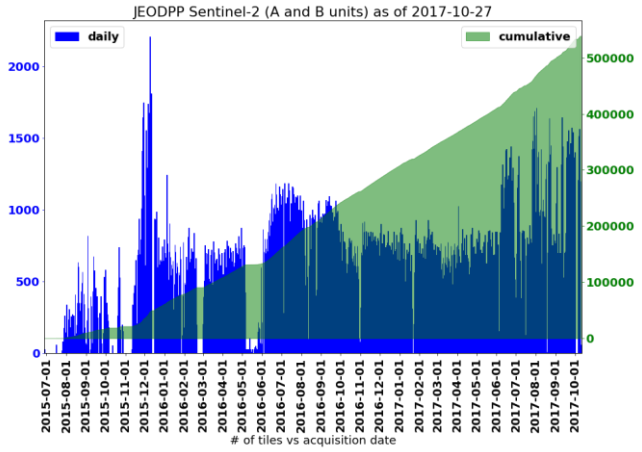

Fig. 2: Number of Sentinel-2 image tiles available on the JEODPP versus acquisition date.

Filesystem in Userspace (FUSE) client. This client acts as a translation layer between a POSIX compliant file system and the native XRootD protocol [4]. This means that all files on EOS storage can be accessed as if they were mounted on a single network file system volume.

3. DATA HOLDINGS

The JEODPP data holdings are driven by JRC user requests.

The bulk of the storage space is currently used by Sentinel1, Sentinel-2, and Landsat data. The data are downloaded from the various data hubs on a daily basis and the data catalogues are updated accordingly. For example, Figure 2 shows the number of Sentinel-2 tiles available on the JEODPP versus their acquisition date. The data are downloaded mainly for Europe and the tropical belt. They correspond to about a fourth of the total amount of available Sentinel-2 data (A and B units). In addition, all Sentinel-2 quicklooks are downloaded to enable application dependent optimal data selection, see example in [7]. Besides Sentinel data, other input data of interest to JRC projects such as Landsat Global Land Survey collections, Meteosat data, and VHR data over selected areas are also included. Further available raster datasets include products such as Global Human Settlement Layers [11], Global Surface Water layers [10], a series of Copernicus Land products as well as several base data such as Digital Elevation Models (EUDEM, 1 and 3 arsec SRTM, ASTER-DEM, etc.), the Copernicus CORE3 2.5m mosaic of Europe [15], a Sentinel-1 global mosaic [18], etc.

Contrary to some geospatial data cube representations [8] where all data sources are converted to a pre-defined coordinate reference system and grid followed by a fixed tiling and stacking along the time dimension, the raster data are stored on the JEODPP in the form of flat files as downloaded from the respective data sources. While pre-computed data cubes provide a faster access to the values of the predefined grid cells along the time dimension, the flat file representation was preferred for its better suitability for general purpose analysis. Indeed, it avoids hard choices regarding the irreversible transformation of the input data to a fixed data cube representation and it is also suitable for heterogeneous data sources including vector data sets without the need to rasterize them.

Besides raster data, the JEODPP also holds a series of vector datasets like European and global administrative units, NATURA 2000 protected areas, EFFIS burnt-area time series, transport networks, etc.

4. JEODPP PROCESSING LAYERS

The three main processing layers are briefly described hereafter.

4.1. Batch processing and containerisation

Scientific workflows developed over the years can be applied to large image collections by distributing the workload to a series of processing nodes. Among the many available workload managers, HTCondor was selected because it is particularly suitable for applications where numerous independent jobs run in parallel without the need for inter-process communication, i.e, High-Throughput Computing (HTC). This is largely the case with satellite images that are often processed in parallel.

To eliminate the problem of installing and administrating application dependent software packages and libraries on the processing nodes and also to avoid the problem of applications having conflicting library requirements, the JEODPP makes heavy use of the light-weight virtualisation based on Docker containerisation [9]. Docker provides an operatingsystem-level virtualisation for flexible management of hardware resources and processing environments by allowing the existence of multiple

isolated user-space instances called containers. In addition, the Docker container-based virtualisation technology integrates smoothly with the HTCondor task scheduler thanks to the so-called HTCondor Docker universe. For applications requiring inter-process communication using for example message passing interface (MPI), HTCondor can be used to submit jobs requiring multiple nodes while the job itself consists of a pool of Docker containers (one per node) managed by Docker SWARM.

4.2. Remote desktop

Some users have developed over the years applications based on dedicated software such as Matlab. Thanks to run-time execution environments, the processing can be launched through batch processing. However, this does not respond to the need for visualising the input or output data in the

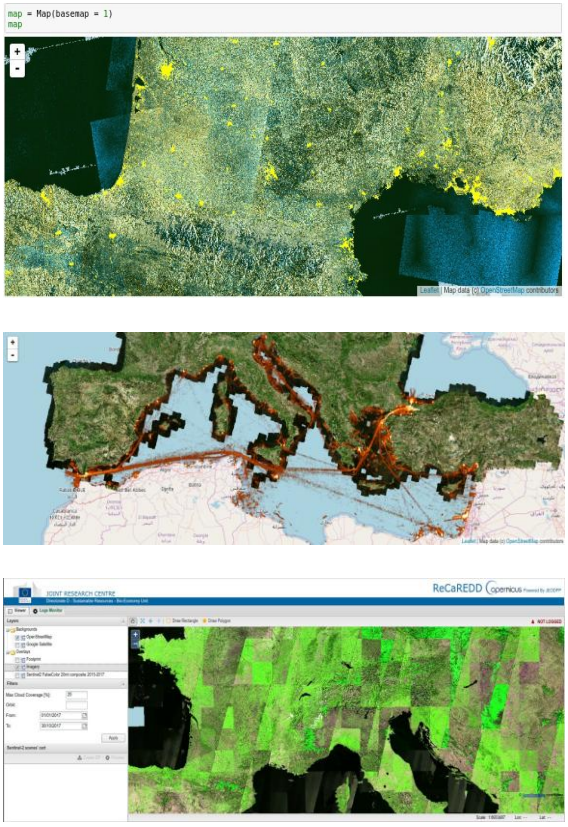

Fig. 3: JEODPP application gallery. Top: Global Human Settlement Layer [2] over Sentinel-1 mosaic. Middle: Ship detection from 2 years of Sentinel-1 imagery over the Mediterranean sea [13]. Bottom: Forest observatory [14].

development environment (besides the interactive visualisation capabilities detailed in Sec. 4.3). This is addressed by offering access to a web-based desktop environment based on Apache Guacamole, a web application that supports graphical access via remote desktop protocols directly in the browser based on HTML5, without the need for additional plugins. Various software libraries and tools such as GRASS, QGIS, and R used by the different JRC projects are provided.

4.3. Interactive visualisation and analysis

The JEODPP also offers the possibility to interact directly with the image data. This is achieved through a web interface that triggers the launch of a Jupyter Python notebook. A dedicated C++ library with Python bindings developed by the project offers the possibility to select and filter image collections and vector data sets, apply a series of processing steps, and render the resulting outputs in a map view area [3, 16]. The processing associated with the rendering is deferred in the sense that it only occurs when the tiles covering the map view area are requested, similarly to the approach followed by the Google Earth Engine [5]. Available

Institutional Platforms

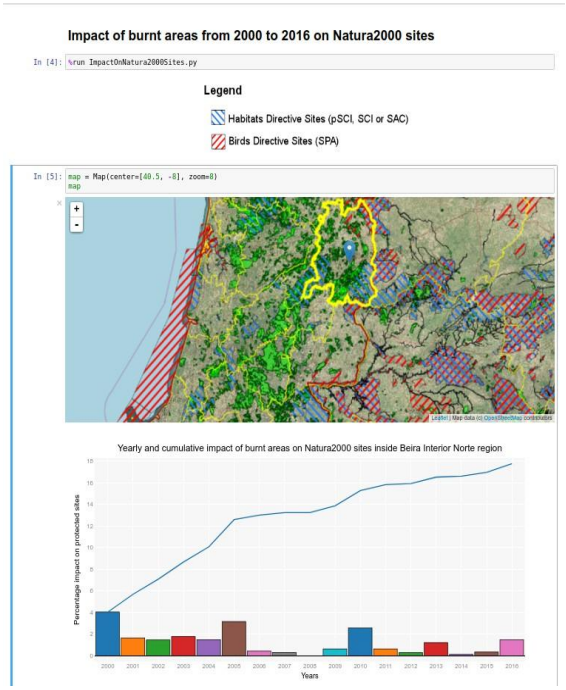

Fig. 4: Example of JEODPP Jupyter notebook showing the interactive computation of the NATURA2000 surface areas affected by forest fires over time within a user-selected territorial unit (a NUTS-3 region in Portugal in this example).

transformations range from simple band combinations to complex connected component based segmentation. Besides users with Python programming skills, targeted information can be conveyed to non-technical stakeholders by simply providing an URL running a notebook with the Python code replaced by widgets, see example with the rendering of digital elevation data in [3].

5. APPLICATION GALLERY

Examples of applications running on the JEODPP are illustrated in Fig. 3: Global Human Settlement Layer [2], the Sentinel-2 web platform for browsing and processing Sentinel-2 imagery for forest cover monitoring over the tropics [14], and Sentinel-1 ship detection [13] using the JRC open-source SUMO software [6]. Besides the interactive visualisation, the Global Human Settlement Layer and ship detection applications are also relying on the batch processing layer for the massive computations involved. A Jupyter notebook with the interactive computation of the impact forest fires on NATURA 2000 sites is shown in Fig. 4.

6. CONCLUDING REMARKS AND OUTLOOK

Data-intensive computing for information retrieval from big geospatial data has recently emerged as a very active field given the availability of massive amounts of free and open Institutional Platforms

geospatial data. The proposed petabyte-scale platform enables the information extraction from large image data sets by users originating from different application domains with their specific data and software requirements. In addition, it contributes largely to knowledge sharing and collaborative working among users with very different levels of computer skills. The project is currently extending to other data sources such as social sensing in collaboration with the activities of the European Media Monitoring[1] by exploiting the geolocation data associated with the collected news and social media items.

7. REFERENCES

[1] Adde, G. et al. “Latest evolution of EOS filesystem”. *Journal of Physics: Conference Series* 608 (2015). DOI: 10.1088/1742-6596/608/1/012009.

[2] Corbane, C. et al. *Global Mapping of Human settlements with Sentinel-1 and Sentinel-2 data: Recent developments in the Global Human Settlement Layer*. Slides of presentation at WorldCover’2017, ESA, Frascati, Italy. Mar. 2017. URL: http://worldcover2017.esa.int/files/2.2p1.pdf.

[3] De Marchi, D., Burger, A., Kempeneers, P., and Soille, P. “Interactive visualisation and analysis of geospatial data with Jupyter”. In: *Proc. of the BiDS’17*. 2017, pp. 71–74.

[4] Dorigo, A., Elmer, P., Furano, F., and Hanushevsky, A. “XRootD — A highly scalable architecture for data access”. *WSEAS Transactions on Computer Science* 4.4 (Apr. 2005), pp. 348–353. URL: http://www.researchgate.net/publication/234817900_XROOTDTXNetFile_a_highly_scalable_architecture_for_data_access_in_the_ROOT_environment.

[5] Gorelick, N. et al. “Google Earth Engine: Planetaryscale geospatial analysis for everyone”. *Remote Sensing of Environment* (2017). DOI: 10.1016/j.rse. 2017.06.031.

[6] Greidanus, H., Thoorens, F.-X., Kourti, N., and Argentieri, P. “The SUMO Ship Detector Algorithm for Satellite Radar Images”. *Remote Sensing* 9.3 (2017), p. 246. DOI: 10.3390/rs9030246.

[7] Kempeneers, P. and Soille, P. “Optimising Sentinel-2 image selection in a big data context”. In: *Proc. of the BiDS’17*. 2017, pp. 177–180.

[8] Lewis, A. et al. “The Australian Geoscience Data Cube: Foundations and lessons learned”. *Remote Sensing of Environment* (2017). DOI: 10.1016/j.rse. 2017.03.015.

[9] Merkel, D. “Docker: Lightweight Linux Containers for Consistent Development and Deployment”. *Linux J.* 2014.239 (Mar. 2014). URL: http://dl.acm.org/citation.cfm?id=2600239.2600241.

[10] Pekel, J.-F., Cottam, A., Gorelick, N., and Belward, A. “High-resolution mapping of global surface water and its long-term changes”.

[1]

Nature 540.7633 (2016), pp. 418–422. DOI: 10.1038/nature20584.

[11] Pesaresi, M. et al. *Operating procedure for the production of the Global Human Settlement Layer from Landsat data of the epochs 1975, 1990, 2000, and 2014*. Tech. rep. EUR 27741 EN. Joint Research Centre of the European Commission, 2016. DOI: 10.2788/ 253582.

[12] Peters, A., Sindrilaru, E., and Adde, G. "EOS as the present and future solution for data storage at CERN". *Journal of Physics: Conference Series* 664 (2015). DOI: 10.1088/1742-6596/664/4/042042.

[13] Santamaria, C. et al. "Mass processing of Sentinel1 images for maritime surveillance". *Remote Sensing* 9.7 (2017), pp. 678/1–678/20. DOI: 10.3390/ rs9070678.

[14] Simonetti, D. et al. *Sentinel-2 Web platform for REDD+ monitoring. Online web platform for browsing and processing Sentinel-2 data for forest cover monitoring over the Tropics*. JRC Technical Report. Joint Research Centre of the European Commission, 2017. DOI: 10.2760/790249.

[15] Soille, P. "Seamless Mosaicing of Very High Resolution Satellite Data at Continental Scale". In: *Proc. of the 2014 Conference on Big Data from Space (BiDS'14)*. Nov. 2014, pp. 222–223. DOI: 10.2788/ 1823. URL: http://publications.jrc.ec.europa.eu/repository/bitstream/ JRC92135/soille2014bids.pdf.

[16] Soille, P. et al. "A Versatile Data-Intensive Computing Platform for Information Retrieval from Big Geospatial Data". *Future Generation of Computer Systems* (2017). DOI: 10.1016/j.future.2017.11. 007.

[17] Soille, P. et al. "Towards a JRC Earth Observation Data and Processing Platform". In: *Proc. of the BiDS'16*. 2016, pp. 65–68. URL: http://publications.jrc.ec.europa.eu/repository/bitstream/ JRC98089/soille-etal2016bids.pdf.

[18] Syrris, V., Corbane, C., and Soille, P. "A global mosaic from Copernicus Sentinel-1 data". In: *Proc. of the BiDS'17*. 2017, pp. 268–271.