

VITO Earth Observation Long Term Data Preservation

Paepen Martine ⁽¹⁾, Vereecken Johan ⁽¹⁾, Ooms Bart ⁽¹⁾

⁽¹⁾ *VITO, Flemish Institute for Technological Research
Boeretang 200, 2400 Mol, Belgium*

Email: martine.paepen@vito.be

ABSTRACT

VITO is using the CVB Archiving Facility (CAF) for the long term data preservation of spaceborne and airborne hyper spectral EO data. The CAF was developed as a modular, generic and portable Java application extendible with a variety of dynamically pluggable modules and is independent from the selected database and storage facilities. The multi mission archiving facility is integrated in the VITO processing infrastructure next to other subsystems like the Data Ingestion Facility (DIF), the Processing Facility (PF) and the Product Distribution Facility (PDF). As long term archive, the CAF is a reliable and flexible operational archiving system which is constantly evolving to meet new business requirements and to adapt to rapidly changing archiving and storage technologies. With the future challenges of the exponential growth of the EO data to be archived and the increasing demand for data retrieval, the CAF is evolving into an archiving system that utilizes HSM (Hierarchical Storage Management) software and which is integrated in a Tiered Storage Architecture.

Keywords: CVB, CAF, HSM, SPOT-VEGETATION.

INTRODUCTION

Since 1998 VITO is hosting the CTIV which is the image processing, archiving and dissemination centre for the spaceborne SPOT-VEGETATION [1] data. Within CTIV, the role of the archive is to preserve the data and to provide data on request, more specifically near real time provision or requests for long time series of a specific region of interest. The operational CTIV archive was designed as a dedicated archive for the spaceborne data received from the SPOT-VEGETATION sensors.

Additional to the dedicated CTIV archive, VITO operates a multi mission archiving facility to archive also higher level products from SPOT-VEGETATION and Earth Observation data from complementary satellite instruments like MERIS, MetOp-AVHRR, AATSR and MODIS. This multi mission Archiving Facility will also act as the Long Term Data Archive for the future PROBA-V user segment.

The multi mission archiving facility, also referred to as the CAF (= CVB Archiving Facility) is integrated in the VITO processing facility where the data can be used in multi mission Earth Observation applications for several environmental research projects. In this way the CAF has become one of the subsystems of the VITO processing and archiving centre next to other subsystems like the Data Ingestion Facility (DIF), the Processing Facility (PF) and the Product Distribution Facility (PDF). The middleware for the communication between the subsystems is implemented by means of a Central Database (CDB). The integration of the CAF with the Product Distribution Facility ensures that the archived data is discoverable and accessible according to standards proposed by INSPIRE and/or GMES.

For the long term preservation of valuable Earth Observation data, the CAF archives two copies on different versions of LTO-tapes. One of the copies is stored at a remote location in the framework of Business Continuity Management. The EMC Networker tool manages, as a plug in on the CAF Software,

the storage on tape. In order to improve its storage management capability the CAF is evolving into an archiving system that utilises commercial HSM (Hierarchical Storage Management) software, integrated in a Tiered Storage Architecture.

CVB ARCHIVING FACILITY AS OPERATIONAL VITO LTA

The CVB Archiving Facility has been developed as one of the sub-systems of the ESA Flexsys project. The goal of the Flexsys project was to define and build a new infrastructure for the development and deployment of new Earth Observation services at VITO in a cost-effective way. The Flexsys software is designed with four main concepts in mind: multi mission, harmonization, interoperability and modularity at all levels. The system integrates independent sub-systems like the Processing Facility, Product Distribution Facility, Archiving Facility, etc. which are loosely coupled. Since the Flexsys project, the Archiving Facility has evolved into the CAF which is a fully operational and independent multi mission archiving facility integrated in the CVB processing and distribution facility.

Integration in the VITO Processing and Archiving Centre

Figure 1 shows the integration of the CVB Archiving Facility as the Long Term Data Archive (LTDA) subsystem of the VITO Processing and Archiving Centre. As such, the Archiving Facility is integrated next to other subsystems like the Data Ingestion Facility (DIF), the Processing Facility (PF) and the Product Distribution Facility (PDF). The middleware for the communication between the subsystems is implemented by means of a Central Database (CDB). The CDB

- registers all incoming products (incl. image data and ancillary data) and all processed products that have to be accessible by other subsystems;
- maintains the metadata, the processing level and the validation status of all products and the relationship of different products with respect to each other;
- handles all the service-specific requests.

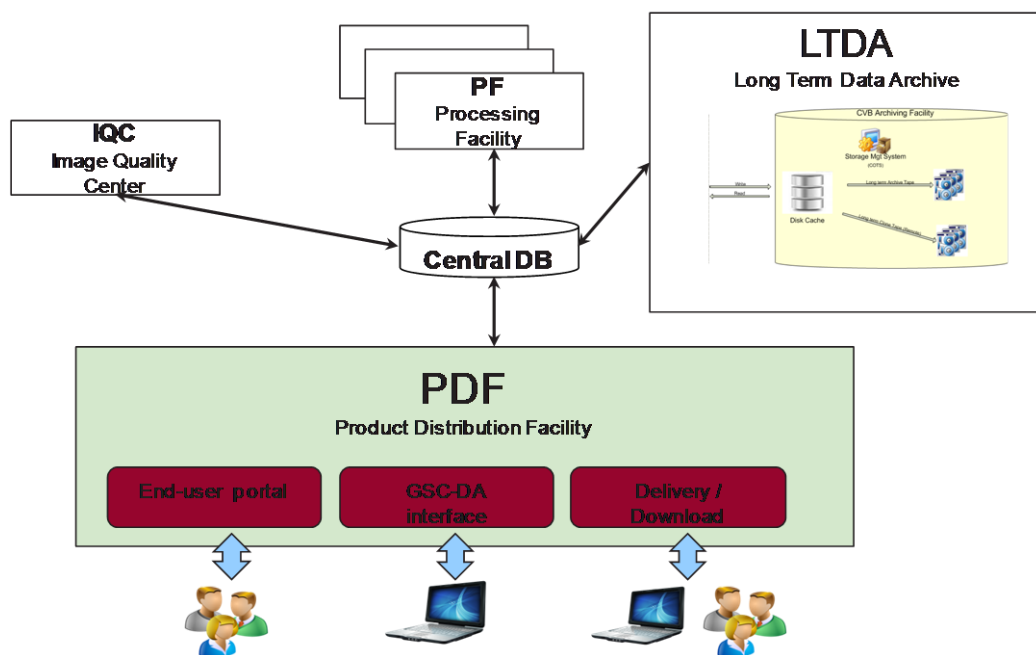


Figure 1: Integration of the LTDA in the VITO Processing and Archiving Centre

Through the CDB, the Archiving Facility is integrated with the Product Distribution Facility to make the archived data discoverable and accessible by means of an interactive web based portal and according to standards proposed by INSPIRE and/or the GMES Space Component. VITO also integrates the historical archive of CTIV in the CDB to create a more standardized and up-to-date way to archive, access and distribute the VEGETATION products for future environmental research.

LTA for PROBA-V products

The lifetime of the SPOT-VEGETATION sensors on board of SPOT4 and SPOT5 is not ensured after 2013. Afterwards, PROBA-V will ensure the continuation of low resolution (1km) Earth Observation products and will provide in addition daily global products at 1/3 km spatial resolution [2]. PROBA-V is a new Belgian satellite, built under the authority of ESA which will be launched in the spring of 2013.

VITO is prime contractor for the development of the PROBA-V User Segment. The main objectives of the PROBA-V User Segment are :

- processing the PROBA-V Level 0 and ancillary data as received from the Secondary Ground Station up to level 3 with a spatial resolution of both 300m and 1 km (for continuity with the VEGETATION data);
- archiving the received and processed data;
- distributing the various products (both at 300 and at 1 km spatial resolution) to the user community;
- assuring the image quality and high level programming of the VEGETATION instrument operations.

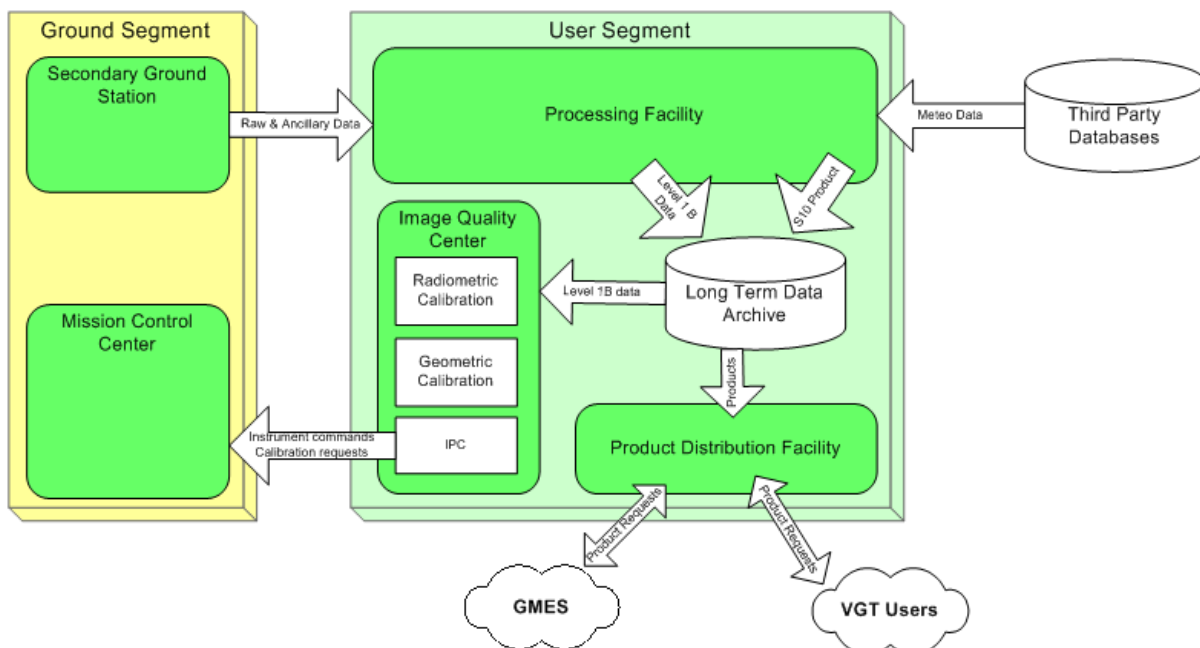


Figure 2: PROBA-V User Segment Blocks and Interfaces

Figure 2 shows a high level diagram with the different blocks and interfaces of the PROBA-V User Segment. For the Long Term Data Archive VITO will reuse the Software Application of the CVB Archiving Facility.

CAF Storage Management

For the long term preservation of valuable Earth Observation data VITO archives two copies of the products on LTO-tapes (one archive tape and one clone tape which is stored on a remote location) and one copy on disk for short term data preservation. The lifecycle specifications to define the storage resource (tape or disk) and the time before being automatically moved to another medium or deleted, can be configured by project via the configuration file. The EMC Networker Software manages the storage on tape.

Figure 3 shows the current CAF Storage Architecture. The CVB archive server hosts the CVB Archiving Facility Software and the EMC Networker Software to archive the data from the EVA Buffer storage to the tapes. The implemented network protocol is a Fiber Channel Network with SAN switches. Our main tape library is a HP ESL322 library with 4 LTO5 tape drives. VITO hosts another robot at a remote location which is also an HP ESL322 library with 2 LTO4 tape drives and is mainly used to generate the second copy of the archived object.

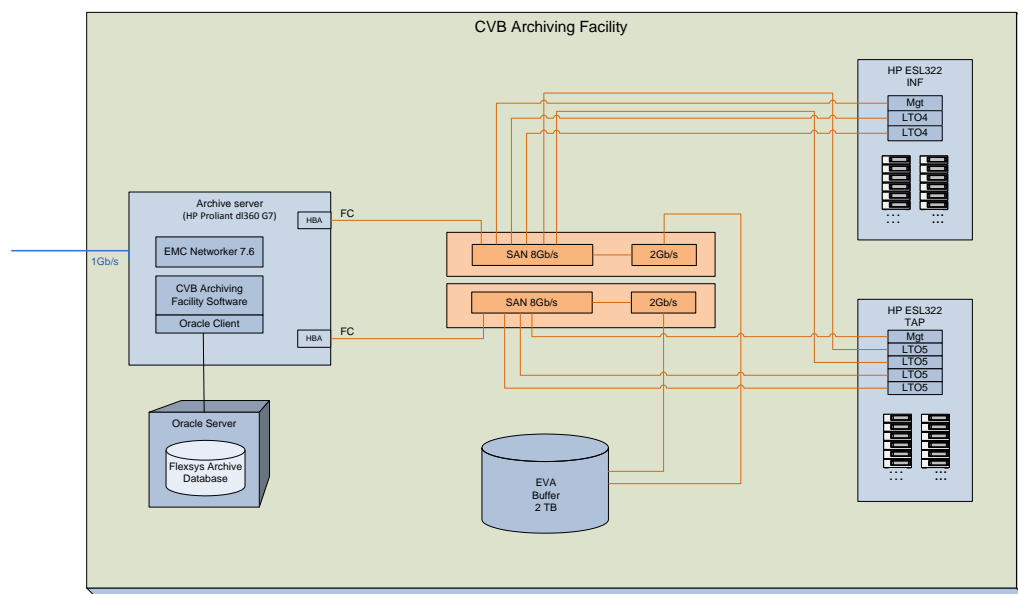


Figure 3: Current CAF Storage Architecture

The CAF software will be reused for the PROBA-V Long Term Data Archive at VITO and the configuration (disk cache / tape libraries / network ...) will be tuned to meet the PROBA-V LTA requirements and data volumes.

EVOLVING FROM CAF SOFTWARE TO HSM

The CVB Archiving Facility is developed using custom made software which includes inter alia storage management functionality. Today there are many reliable and mature COTS available to ensure the storage management in a reliable, secure and perform way. To reduce the CAF maintenance costs and to face the future challenges of the exponential growth of the EO data to be archived in the LTA and the increasing demand for data retrieval, the CAF will evolve towards an archiving system that utilizes COTS HSM (Hierarchical Storage Management) software and which is integrated in a Tiered Storage Architecture.

The high level diagram of the current CAF configuration and integration with the central DB (see Figure 4) shows that the core of the archiving facility is a relational database that is used for both:

- the metadata repository (inventory)
- and the persistent layer of the data management of the archiving facility.

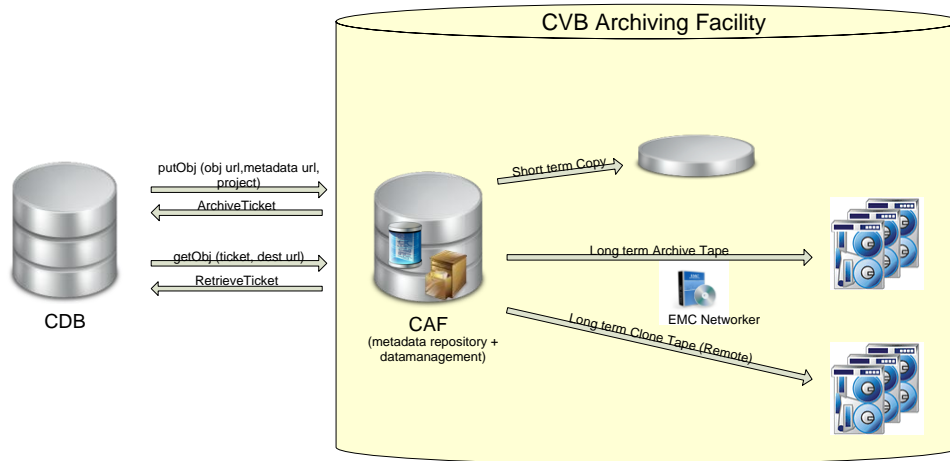


Figure 4: High level CAF Configuration and Integration with CDB

The data lifecycle specifications to define the storage resource (tape or disk) and the time before being automatically moved to another medium or deleted, can be configured by data type via the configuration file. The CAF database stores all the information about the data management (e.g. storage resource, media volume name, access URL, migration date, etc.). This information is used to retrieve the archived data and also by the maintenance job which is scheduled on regular (configurable) times, and triggers the data migration and deletion based on the information in the CAF database. Additionally the EMC Networker Software is used as a plug-in on the archive software to store the data on LTO tape.

The main triggers for the CAF optimizations by using COTS HSM Software are:

- elimination of EMC Networker Software because it is evolving more and more into a real back-up tool with less focus on archiving;
- disconnect the storage management functionality from the CAF software in order to reduce the maintenance and support effort/costs;
- elimination of the different plug-ins for disk or tape storage;
- the feasibility to integrate the archive in a Tiered Storage Architecture.

Figure 5 shows the CAF optimizations by using HSM Software.

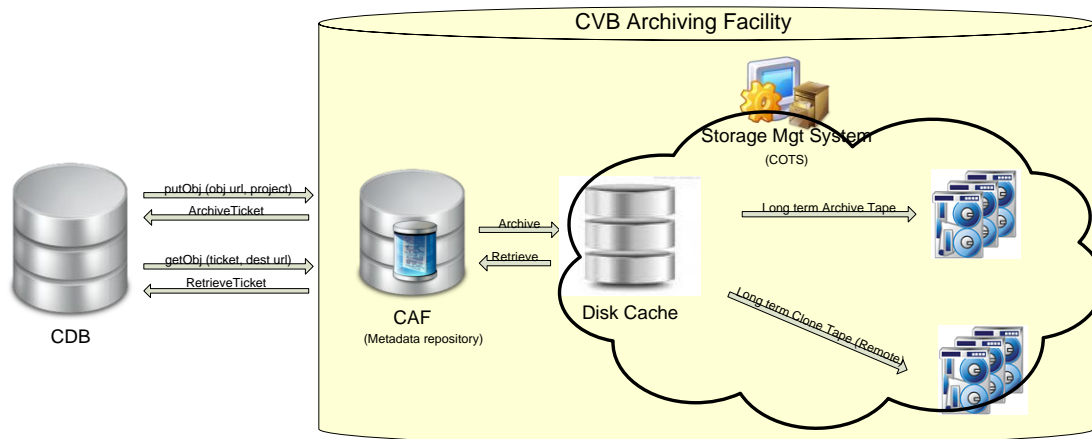


Figure 5: CAF optimizations by using HSM Software

The interface to the current CAF shall be retained and also the metadata repository which is an inventory on the archived products. The COTS storage management system will cover the data storage and management functionality on different media (e.g. tape and disk) and will provide reports and monitoring tools on the storage environments.

CONCLUSION

The CVB Archiving Facility is a reliable and flexible operational archiving system which is constantly evolving. To face the future challenges of the exponential growth of the EO data to be archived for the long term and the increasing demand for data retrieval, the VITO multi mission EO archive is evolving from the custom made CAF software with the EMC Networker plug-in towards a system that utilizes HSM Software to manage the data following well defined policies and which is integrated in a Tiered Storage Architecture. Since the EO data, stored at VITO, has evolved from mission-critical data only towards also less mission-critical data, the current enterprise class only system architecture involves avoidable costs. With a Tiered Storage Architecture we aim to increase storage efficiency, to accelerate data retrieval, to improve data management and to reduce overall costs. Hereby the goal is to enhance the alignment between data business value and the provided service level.

REFERENCES

- [1] – SPOT-VEGETATION: <http://www.spot-vegetation.com/>
- [2] – PROBA-V: http://www.esa.int/SPECIALS/Proba/SEMD16ZVNUF_0.html