# Swarm Payload Data Ground Segment: Status, Evolution and Lessons Learned

**L. D'Alba** [1], A. de la Fuente [2], A. Maltese [3], D. Parente [3], V. Panebianco [3], L. Mariani [3], A. Biscuso [3], F. Gabrieli [3]

[1] RHEA c/o ESA [2] ESA-ESRIN - [3] SERCO c/o ESA

## Introduction


The **Swarm Payload Data Ground Segment (PDGS) is the mission's element in charge of the data processing, archiving, and dissemination; products quality control; calibration and performance monitoring**. The Swarm PDGS functionalities are implemented as a **combination of dedicated systems and teams, and shared services**. It has been **performing without** major **problems** since its inception fulfilling all the mission's requirements and representing a major success for the Swarm mission thanks to a continuous evolution.

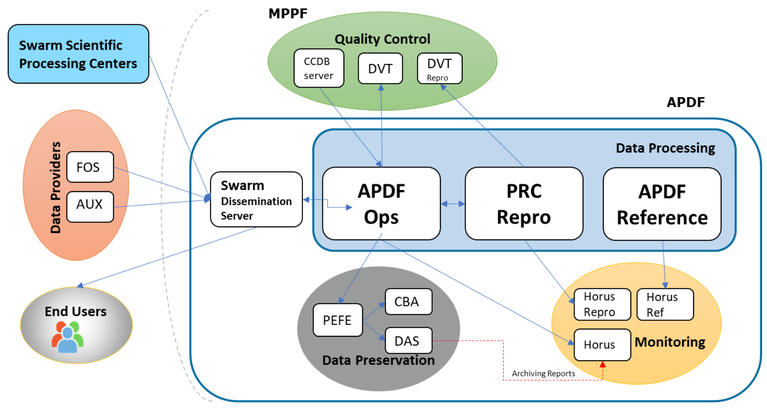
The **architecture of the Swarm PDGS is always evolving** to respond to new operational and science requirements, the incorporation of new products, to improve the robustness, maintainability, and efficiency of the current system using the latest available techniques. Each single system and the entire infrastructure are continuously upgraded to be compliant with new security policies.

This poster summarises the **efforts of the Swarm PDGS team during the past decade** in several areas: **Swarm PDGS functions and status, APDF infrastructure migrations, reprocessing activities, data access, main achievements and lessons learned.**

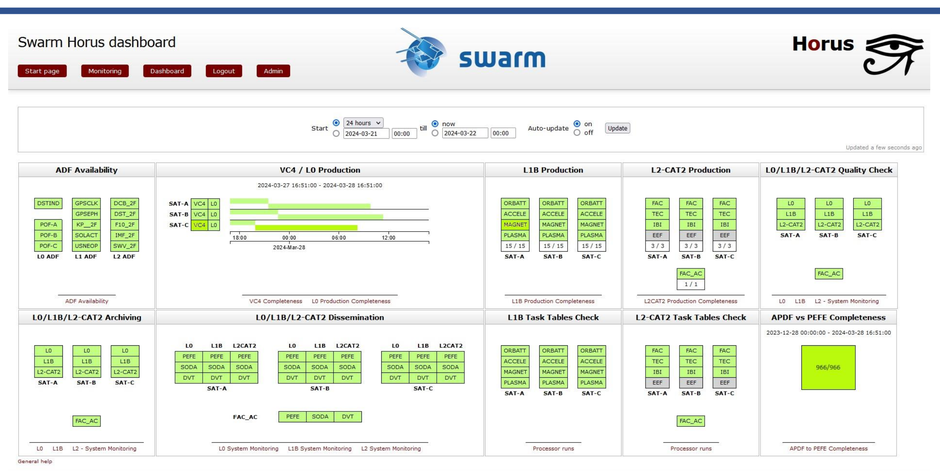
## Swarm PDGS Functions and Operational Status

The main **Swarm Payload Data Ground Segment** (PDGS) functions are **distributed** among different subsystems/places:

- The **Archiving, Processing and Dissemination Facility**, **APDF**, is made up of the **Data Processing element** (including the Operational and the Reprocessing Chains and the Reference Platform) hosted at ESRIN; the **Swarm Dissemination Server**; and the **Data Preservation element** implemented using an **EOP-GE common service**. The **E2E monitoring** is done by the **Horus** tool.
- The **Mission Performance and Planning Facility** (MPPF) in charge of the **end-to-end performance monitoring and quality assurance** is also located at ESRIN.

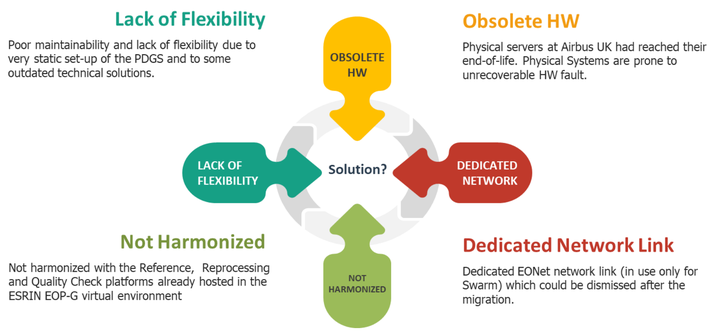



The Swarm **Scientific Processing Centers** (**SPCs**), responsible for the generation of L2-Cat1 products, are distributed across various European and American institutions. These centers process the data and subsequently relay it to the APDF for archiving and dissemination.

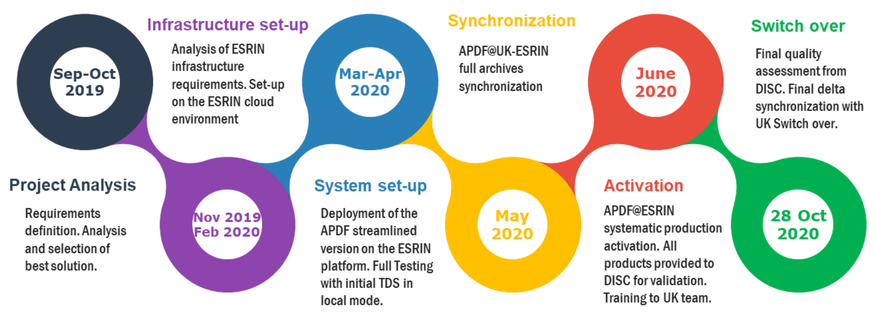



**Operational Status:** All the PDGS functions have been **running nominally** during the last year. The **Horus Dashboard, E2E monitoring system which has been strongly evolved during the mission,** shows the status of ADF/VC4 ingestion, the L0/L1B/L2-Cat2 production, quality checks, dissemination and archiving functions over a selected time period or in real-time.

## APDF Platform Migrations to new Infrastructures: Past evolutions and Future plans




**Migration Rationale:** Several years ago, it became apparent that the **Airbus UK's APDF platform was unsustainable** in the medium to long term due to its **physical servers reaching end-of-life (EoL) and an obsolete PDGS architecture**. Upon thorough evaluation, the decision was made to **transition to ESRIN's EOP-G virtual infrastructure**, thereby phasing out the obsolete dedicated network link.




**Migration Timeline:** The **APDF migration** has been a **complex project** spanning **several years**. After the set-up of the new platform on the ESRIN cloud environment where the Reference and Parallel Reprocessing chains was already deployed, the **nominal operations were running in parallel for several months**. After **an extensive validation** of the ESRIN production, **the platform switch-over** was activated on **28 Oct. 2020**, and the UK infrastructure was decommissioned. The migration has included an **optimization** of the **internal dataflows** and a **configuration fine tuning**.

The PDGS OPS Team is initiating a **new migration** to the **ESRIN EDGE POD** infrastructure, expected to conclude by the end of Q2 2024. This transition will be seamless to the user community and will not affect their experience. The primary reason for this update is **to enhance the maintainability** of our existing systems.

## Two Full Mission Reprocessing Campaigns Completed


**Full Mission Reprocessing Campaigns** are **complex activities** involving several teams (SW maintainers, PDGS Ops team, Data Quality experts, Scientific Community, etc.) which require a **careful planning**. The aim of the overall process is to provide the end users **fully consistent product datasets** with the **best possible quality**, including the latest improvements in the algorithms and calibrations.

The **Swarm full mission reprocessing campaigns** are carried out on an independent **Swarm Parallel Reprocessing Chain**, a **streamlined, cloud-based and scalable** version of the operational chain, hosted at the HCINSS ESRIN Virtual Infrastructure and specifically optimised in order to **maximize** the **processing throughput** providing **flexible provision of the required storage and processing power**.

The current infrastructure boasts a **throughput capacity** of processing approximately **one year's worth of data from three satellites** in about **three days for Level 1b** and in around **one day for Level 2-Cat2**. The **1st Full Mission Reprocessing concluded in 2018, followed by the 2nd in 2022**.

Advanced access to the final reprocessed datasets is granted to Quality Expert teams for validation. Once verified, these datasets are integrated offline into the operational chain and subsequently distributed to all users via the Swarm dissemination server. Concurrently, the new processing baseline is implemented in the Nominal Chain.

It is expected to start the **3rd Reprocessing Campaign in the upcoming months**. This campaign aims to incorporate all the planned enhancements for the processor, which are due for release in L1b version 3.25.

## Swarm Data Access

Products from the Swarm mission are promptly accessible for Mission Teams and Regular Users via the **Swarm Dissemination Server**, which supports **FTP, FTPS, and HTTPS** protocols at the address **swarm-diss.eo.esa.int**.

This server also hosts **Swarm advanced datasets**, a mirror of the **CASSIOPE/e-POP** full data archive, and **Cryosat-2 magnetic data**.

Since the end of Commissioning about **43.2 Million of files (490 TB) have been distributed** via the dissemination server, in particular **7.5 Million** (**72 TB**) during **the last year.**

**Steady growth of the Swarm data user community** with **access peaks** associated **Conferences, Workshops** and publication of new product updates or **Reprocessing Campaigns**.

**Update on Data Access Strategy:** The ESA Single Sign-On system will soon become the exclusive method for accessing the Swarm Dissemination Server and downloading data. This measure is intended to streamline access and enhance security by phasing out all functional accounts in favor of personal accounts only. A gradual notification process for all users is underway to ensure a seamless transition.

## Conclusions and Lessons Learned

**The Swarm PDGS is performing exceptionally well,** a testament to the effective synergy of specialized systems, dedicated teams, and shared services. Notable accomplishments over the previous ten years include:

- **Refinement and optimization** of the system's architecture for sustainable long-term maintenance.
- Fortification and **improvement of system security** measures.
- Transitioning from **physical to virtual infrastructure**, with forthcoming enhancements.
- **Streamlined and efficient** methods for the user community to **access data**.
- Rollout of the **FAST chain**.
- **Enhancements to the monitoring and reporting** mechanisms, boosting functionality and reducing security risks.
- Incorporation of **new data products from DISC** scientific processing centers and supplementary missions.
- **Revision of contractual strategies** to emphasize the integration of maintenance, evolution, and primary level operations.

**Main lessons learned:**

- The Swarm PDGS infrastructure has been **continually adapting** to meet new operational and scientific demands, integrate new scientific outputs, and enhance its robustness, and efficiency by employing state-of-the-art methodologies. Cost efficiency is achieved through collaborative efforts, adaptability, ongoing development, streamlining processes, and the implementation of Agile methodologies.