Ten Billion Transfers per Day and How to Follow Them – The Evolution of the System Monitoring and Reporting Tools

Víctor Sierra Urueña[a], Tristan J. Edwards[b]

[a] *System Operations Engineer, Real-time Services and System Operations Division (RSO), European Organisation for the Exploitation of European Satellites (EUMETSAT), EUMETSAT Allee-1, 64295 Darmstadt, Germany, victor.sierra@eumetsat.int*
[b] *System Operations Team Leader, Real-time Services and System Operations Division (RSO), European Organisation for the Exploitation of European Satellites (EUMETSAT), EUMETSAT Allee-1, 64295 Darmstadt, Germany, tristan.edwards@eumetsat.int*

Abstract

In the past 30 years, EUMETSAT has expanded its portfolio of operational products and services exponentially in the role of becoming a key International Meteorological Data Exchange entity. Over a quarter of a million scientific products per day are now disseminated on EUMETSAT's prime delivery mechanism, EUMETCast. These are constituted not only of data from satellites operated by EUMETSAT but also from a diverse array of operational partners and agencies supplying data for retransmission to our users. Complementing this is an expansive archive of products allowing users direct access.

Behind the scenes of these delivery mechanisms lies a highly complex ground segment architecture, which provides a complete end-to-end service including; raw data reception, higher level processing and re-distribution to users through various interfaces and mechanisms. The total number of movements in these products around the ground segment has now reached the milestone of over ten billion per day, and is set to take another quantum leap with the introduction of the new MTG and EPS-SG mandatory programmes and their associated products.

One major challenge therefore is to what extent these transitions can and should be monitored, and how. EUMETSAT's response, developed as a proprietary software over the past 15 years, is the Service Monitoring and Analysist Tool called SMART. The role of SMART is to interpret system log files for events pertaining to product processing and dissemination and compare them against a schedule of expectations. The resulting comparison of "actual vs expected" is key to facilitate a quick system overview but also in other areas such as longer term Key Performance Indicator (KPI) reporting or providing a system status to our users via the EUMETSAT website.

This paper reviews the current status of SMART, including its strengths and challenges, as well as the concepts for developing it in the future in combination with a machine learning framework to facilitate the growth in number of products foreseen at EUMETSAT. The introduction of machine learning and automation are seen as a critical step in facilitating the monitoring without a significant overhead being spent in configuration. The paper also looks at SMART in the overall concept of system monitoring and the related challenges of finding the correct balance in the level of monitoring of individual ground segments and multi mission elements.

Keywords: EUMETSAT, Machine Learning, Monitoring, Reporting, KPIs.

Acronyms/Abbreviations

ADM	Atmospheric Dynamics Mission
CNES	Centre national d'études spatiales
COTS	Commercial off-the-shelf
CPF	Common Processing Facility
DLR	Deutsche Zentrum für Luft- und Raumfahrt (German Space Agency)
ECast	EUMETCast
EPS	EUMETSAT Polar System

EPS-SG	EUMETSAT Polar System – Second Generation
ESA	European Space Organisation
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
GTS	Global Telecommunication System (of the WMO)
IJPS	Initial Joint Polar System
JMS	Java Message Service
KPI	Key Performance Indicator
MCC	Mission Control Centre
MMDS	Multi Mission Dissemination System
MME	Multi Mission Element
MPT	Mission Performance Tool
MSG	Meteosat Second Generation
MTG	Meteosat Third Generation
NASA	National Aeronautics and Space Administration
NOAA	National Oceanic and Atmospheric Administration
NLP	Natural Language Processing
OS	Operating System
RMDCN	Regional Meteorological Data Communication Network
RRD	Reduced Resolution Dataset
WMO	World Meteorological Organization (UN)

1. Introduction to EUMETSAT

EUMETSAT is an intergovernmental organisation based in Darmstadt (Germany), responsible for the exploitation of Europe’s meteorological satellites.

EUMETSAT operates a system of meteorological satellites that observe the atmosphere, ocean and land surfaces – 24 hours a day, 365 days a year. This data is then supplied to the National Meteorological Services of the organisation's Member and Cooperating States in Europe, as well as other users worldwide.

The satellites currently operated in the EUMETSAT HQ are:

- Geostationary satellite MTG-I1 (Meteosat Third Generation – Imaging 1) launched in December 2022 and in commissioning phase until end of 2023. It will provide imaging services supporting nowcasting applications.
- Geostationary satellites Meteosat -10, and -11 over Europe and Africa, Meteosat-9 over the Indian Ocean. This corresponds to the Meteosat Second Generation.
- Metop polar-orbiting satellites (Metop-B and Metop-C) as part of the Initial Joint Polar System (IJPS) shared with the US National Oceanic and Atmospheric Administration (NOAA).
- Jason-CS/Sentinel-6 satellite providing global sea surface height observations for climate monitoring and ocean and seasonal forecasts. Jason-3 is also part of the Jason mission (an international partnership between EUMETSAT, CNES, NOAA, NASA and the European Union via the Copernicus programme), even though it is not operated by EUMETSAT.
- Sentinel-3 satellites (S3A and S3B) collecting observations of global ocean colour, sea surface temperature and sea surface height.

One of the main objectives of EUMETSAT is also to create synergies with other operators of Earth observation satellites. Currently, EUMETSAT cooperates with other agencies including in Europe, China, India, Japan, South Korea and the United States, benefiting from the sharing of data from many other satellites.

The data and products from both EUMETSAT and third party organisations are vital to weather forecasting and make a significant contribution to the monitoring of the environment and climate change. They aid meteorologists in identifying and monitoring the climate change or the development of potentially dangerous weather situations that affect air travel, shipping, road traffic, farming, constructions and many other critical industries.

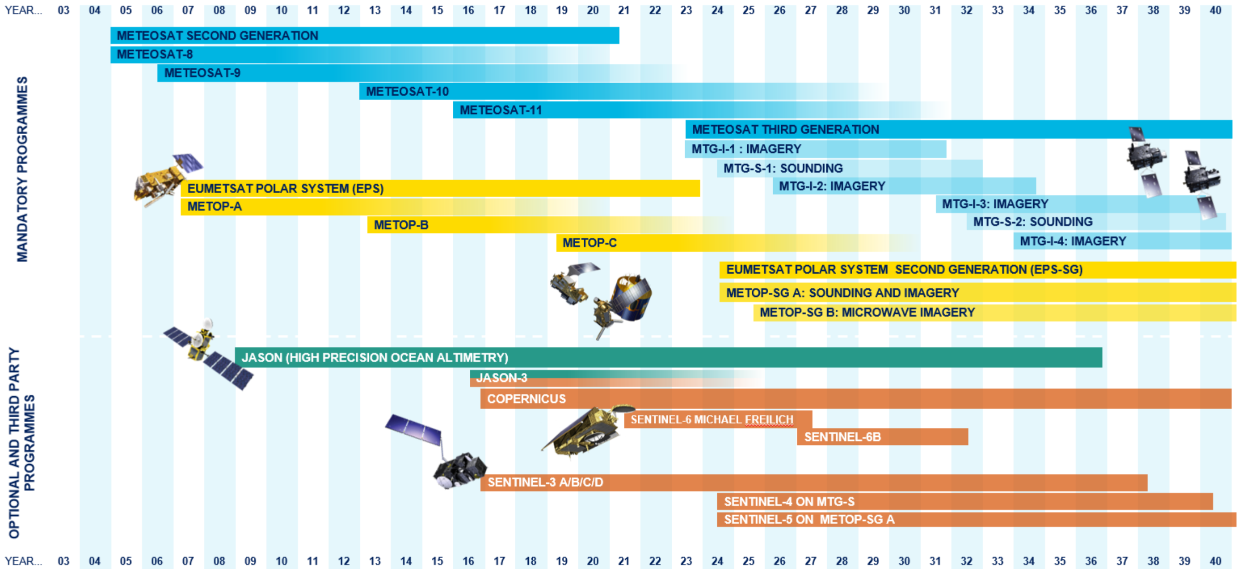

Fig. 1. EUMETSAT Operational Mission Planning 2000 - 2040

As it can be observed in figure 1, EUMETSAT’s planned missions and launches will continue to grow during the next 20 years. This increasing number of satellites and, consequently, amount of data that needs to be handled generate a big impact on the structure of the organisation. The addition of missions requires not only scaling many of the existing systems, but also redesigning those that become inefficient or simply unable of handling the new load.

One of the major challenges facing EUMETSAT: how can such a large number of transfers be monitored in an efficient manner, without incurring in an excessive monetary and human expense? This paper will explain the tools that EUMETSAT adopted to monitor these data transactions and the main challenges and future evolutions expected in the coming years.

2. Mission Control Centre at EUMETSAT

The EUMETSAT Mission Control Centre (MCC) is based at its headquarters in Darmstadt (Germany) and is divided into two control rooms, one for the geostationary (GEO) missions and the other for the Low Earth Orbit (LEO) missions. Shift teams of satellite and ground segment controllers work 24 hours a day, supported by teams of on-call operators and maintenance engineers.

The MCC is the part of the overall ground segment responsible for the safe operation of all satellites. It provides monitoring and control functions for the spacecraft and antennas, but also monitoring of science data (L0, L1, etc.) and supporting infrastructure. It also provides reporting functions in order to notify the user community in case of expected or unexpected events, display the status of each mission in a real-near time manner, or generate KPI reports.

For most of these functionalities, EUMETSAT makes use of Multi-Mission Elements (MMEs), which is any system that can be “recycled” across multiple missions. MMEs can be represented by physical servers, software tools, processes or all at once. Their usage brings clear advantages to the organisation such as reduction of development and maintenance time and cost, and taking advantage of the clear similarities across the space missions. One of the disadvantages of using MMEs is the lack of flexibility when it comes to adapting to programme specific requirements, which is usually ameliorated with plugins or smaller ad-hoc supporting tools.

The monitoring and reporting solutions adopted at EUMETSAT are also considered MMEs, thanks to the fact that all missions share certain functions (such as acquisition, processing, dissemination or archival of mission data).

While there are several tools at EUMETSAT that address these functionalities, this article will focus on two: GEMS and SMART.

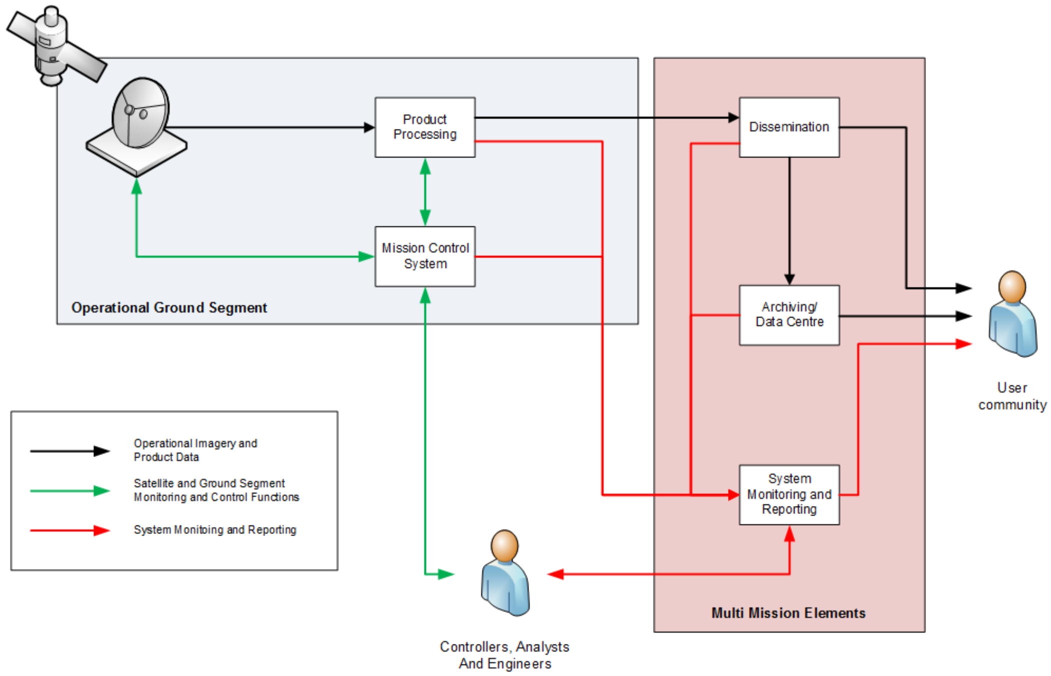

Fig. 2. Mission Monitoring and Reporting in the EUMETSAT context

3. Mission Performance Monitoring and Reporting at EUMETSAT

EUMETSAT delivers services to the user community on a 24/7 basis and therefore it is vital to monitor the infrastructure and operations that support them. The monitoring and reporting of mission data allows a quick response to anomalies and a better characterization of the issue and components affected (for instance, failure in antenna *Y* or misconfiguration in the Level 2 image processor).

EUMETSAT utilizes two proprietary solutions (MMEs) for monitoring and reporting, developed over more than 10 years and whose functions are diverse yet complimentary: collect logs from any operational system (GEMS) and compare existent data flows against user expectations (SMART). In addition, EUMETSAT also uses COTS such as OP5 (based on Nagios).

Overall, these tools allow the organisation to control and understand any past and present event that occurs in any operational system.

3.1. Generic Event Monitoring System (GEMS)

The Generic Event Monitoring System (GEMS) is used to gather, in an organised way, logs from all mission specific and MME systems into a centralised database, with a user interface for near-real time monitoring by controllers and offline queries of historical events.

GEMS divides its functions into a client process and server process. The GEMS client, installed in an operational server, is in charge of reading a particular machine event log (using its own local agent or via an API) and transferring the relevant information to the GEMS server through a selection of protocols such as FTP, HTTP (RESTful) and JMS Message Queue. The GEMS client is a lightweight Java application designed for use on a wide selection of UNIX/Linux and Windows based platforms.

The GEMS server is in charge of receiving these logs and storing them in a Central PostgreSQL Database driven by a Web server application (i.e. accessible via a UI). Events are grouped into logical "facility" devised from the programme, environment (operational, validation) and particular function of the facility (e.g. product processing, dissemination). Events within a facility can consist of either system/OS level events or events from application logs.

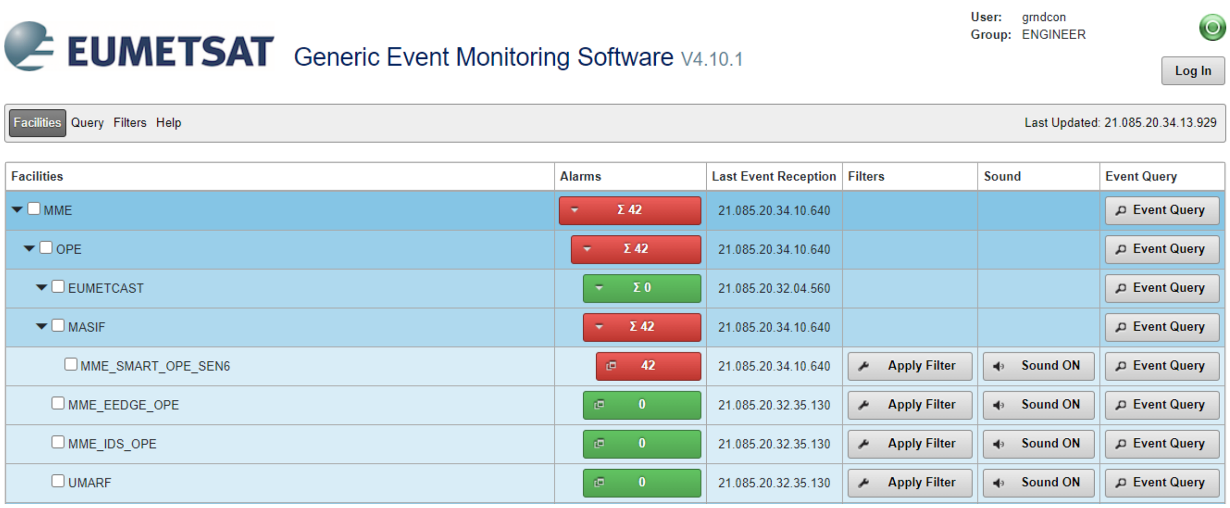

Fig. 3. GEMS graphical user interface

GEMS distinguishes between three types of events: Information, Warning and Alarm. The latter group, by default, generates a particular audible message that can be heard in the Mission Control Room, triggering recovery actions by the person in charge. Information-type events are generally used to indicate the reception or dissemination of a particular satellite product from point A to point B; this information is later used by SMART to perform the end-to-end data flow monitoring, as it will be presented in the following section.

As of January 2023, GEMS processes in the region of 30 million events per day equating to ~10 billion events per year from more than 200 different hosts.

3.2. System Monitoring and Reporting Tool (SMART)

The System Monitoring and Reporting Tool (SMART) represents an additional monitoring layer, which uses GEMS events as its main input. While GEMS stores logs and classifies them by facility or server type, SMART is able to track individual products from the point of product production in the mission-specific ground segments, through the dissemination chain right to the end users.

The main purpose of SMART, however, is not only to track the distribution of products through operational servers, but also to compare this against a pre-configured schedule. Doing so, SMART can alert its users of a missing product, and recognise the point in the dissemination chain where the anomaly most likely occurred.

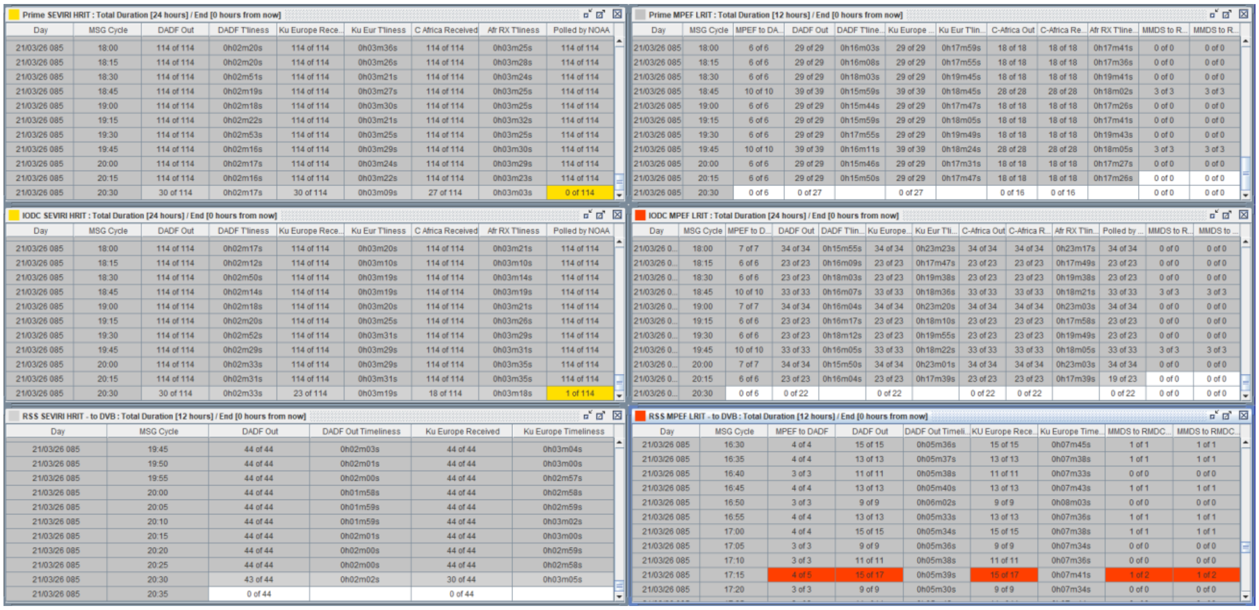

Fig. 4. SMART graphical user interface (MSG viewers)

Not only can this schedule be based on static data, but also on an operations schedule of activities, Mission Planning files, etc, to create a dynamic and accurate set of expectations. From this, SMART can determine the completeness, timeliness and to a certain extent, also quality of all products that form part of the services EUMETSAT provides to its users. While originally developed for detailed near-real time monitoring, SMART is exploited for other uses, including the near-real Operational Service Status Indicator, and powerful offline reporting of KPIs.

3.3. OP5 Periodic Monitoring

OP5 is an enterprise version of the third party Nagios product, which is widely used in industry for periodic monitoring of systems. It performs basic checks on servers, applications and COTS processes (e.g. CPU usage, physical memory, system load, availability status) once every five minutes (on average) and is highly configurable in terms of metrics and their frequency.

When one of the OP5 requests results in a failure, the tool will increase the frequency of the poll so that the system is evaluated more accurately. After a specified failure threshold, OP5 will raise an alarm via UI, email or any other configurable mean (such as GEMS).

OP5 currently monitors in the region of 4000 operational servers providing approximately 70,000 metrics regarding the state of the operational system. In addition, the storage capacity used for holding the performance data is efficient - OP5 is relying on RRD files that can keep for each hosts/service statistics for many years in a less than 1 MB file. As a result, the system is using less than 30 GB disk space for performance data over a 2 year period.

3.4. Monitoring and Support Infrastructure Facility (MASIF)

The MPT server components are hosted on the Monitoring And Support Infrastructure Facility (MASIF), which provides the hardware, COTS, network and storage for the applications that complete the Mission Performance Toolset. MASIF is itself considered an MME as it is not part of any one mission, but centralized to them.

The MASIF infrastructure is built with scalability in mind. Using a virtual machine concept based on VMware, virtual machines can be swapped across physical hosts with complete transparency, and even the entire underlying hardware upgraded without any noticeable impact to the running applications.

4. SMART Architecture

An operational SMART system consists of a SMART Relay (containing the Web Client front-end component through which viewers and reports can be accessed) and one or more SMART Instances (i.e. an independent deployment generally represented by a mission, such as MSG, MTG, S3, S6, etc.).

The SMART configuration is stored in XML and ASCII files while the SMART schedule data is stored in a Relational Database Management Server (RDBMS), such as Oracle 11g, PostgreSQL 11.x or Apache Derby.

SMART functioning follows this sequence:

1. It generates a schedule of events (e.g. 10 events expected at every hour).
2. It checks in the GEMS database and tries to match events using configurable regular expressions.
3. When an event is matched, SMART extracts the key information of the product and writes it into the SMART database.
4. The user can access the corresponding SMART viewer or report, which will show "1 of 10" events in the particular cycle (after the event has been matched correctly). The user will also be able to observe which is the product matched (i.e. its key parameters) and which are still missing.

SMART also supports the creation of annotations, which can be used to note down special conditions such as outages, which apply during a specific time window and are linked to an instance, service or product. Annotations can be either created by users manually, or imported from EUMETSAT announcements.

SMART components are decoupled from each other, i.e. a failure of one component of the system does not cause loss of total monitoring, as it is represented in figure 5.

Error! Reference source not found. Fig. 5. SMART Context Diagram

Inside the block of SMART Configuration (figure 5), SMART utilizes different types of files in order to operate. A summary of the most relevant files and their main purpose is defined in the following sections.

4.1. SMART Configuration files

The SMART configuration files define the structure of a particular service (e.g. MSG 0 Degree Service, MSG IODC Service, etc.) by dividing it into **trackers**. A tracker represents a point in the data chain that needs to be monitored. Examples of tracker are:

- Product Received at Dissemination Facility
- L0 product Generated
- L2 product Generated
- Product Archived
- Product Received at Reference User Station

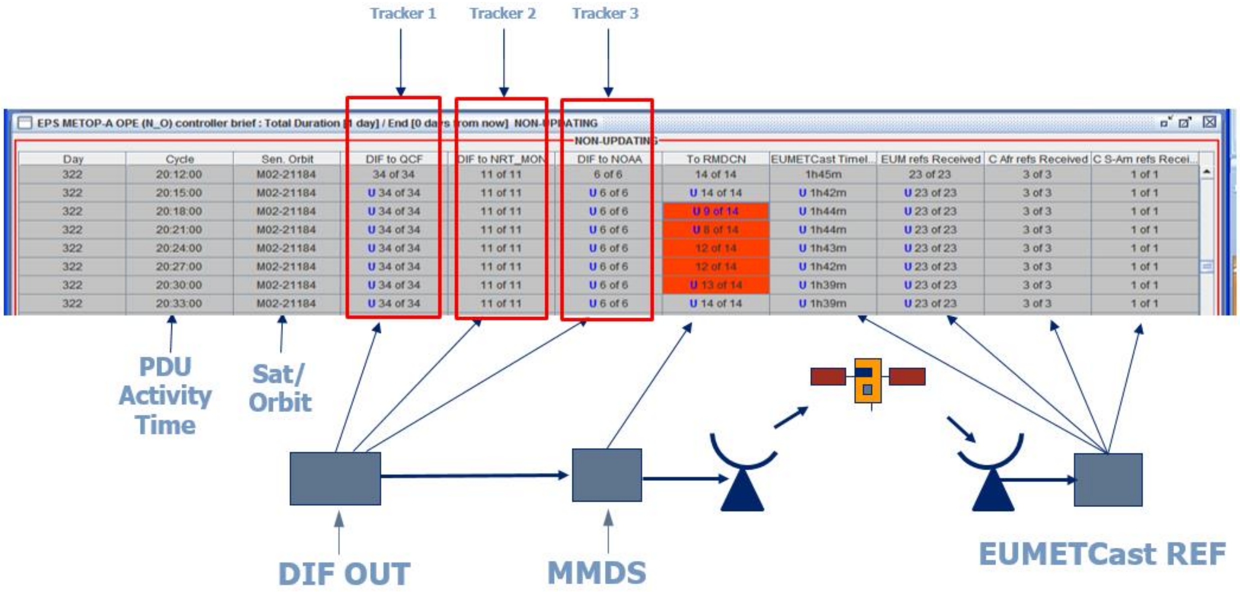

Fig. 6. SMART Viewer example

Each tracker uses configurable regular expressions to filter and match GEMS events. When a GEMS event is found, SMART will extract its key **parameters** (e.g. Timestamp, Host, Process, Sensing Start, Sensing End, File Size, etc.) and store it in its own database server.

4.2. SMART Schedule files

The schedule files define the expectation for each particular product that is found by the configuration files. Therefore, for each product that is planned, there is normally an entry in the schedule indicating its expected key parameters, timeliness and frequency.

SMART is then able to generate a schedule of expected events in the future, that will be compared against the real data and show if it is met or not in terms of:

- **Completeness** → Have all expected products been received (for a particular day and time)?
- **Timeliness** → Have all expected products been received (for a particular day and time) in their due time?

Both completeness and timeliness play an important role in SMART, since they are the main inputs to generate KPI reports.

4.3. SMART Viewer and Report files

The viewer and report files define the information that is displayed in the screen for a specific SMART service. While a configuration file might include e.g. 20 trackers (or points in the data chain), a viewer can display any particular subset.

The viewer and report files are generally configured to create different SMART views focused on spacecraft/ground controllers, engineers or management.

5. SMART and GEMS Limitations and Challenges

SMART and GEMS tools present various challenges that have been considered over their operational lifetime, a selection of which are presented below. In fact, these are also considered in the frame of developing new software and processes.

5.1. Schema-on-write vs Schema-on-read

One known limitation of SMART is the schema-on-write nature of the current software [1]. For GEMS, although metadata is applied (Timestamp, Host, Process and Message text), the content of the messages themselves is free and completely uncategorised. That makes GEMS a schema-on-read software. SMART, on the other hand, extracts further parameters from GEMS events, compares them against set expectations and stores this new schema in the database. Therefore, working as a schema-on-write software instead.

Although this does not create a major problem under nominal conditions, having two databases with different schemas imposes a great limitation when it comes to changing any of the stored data. Suppose that a certain service is created with an expectation of 5 daily products, and after 6 months it is decided to change a parameter of one of them (for example, its “maximum timeliness”). Since SMART has already created a rigid database in the first 6 months, in order to apply the new changes it would be necessary to delete it completely and process all past data again. This would mean, in certain cases, reprocessing some millions of GEMS events and spending several hours of computational work.

5.2. SMART configuration load – new missions

Another major challenge is the increasing time required to modify the SMART configuration, or the impossibility of automating it. As seen in section 4, SMART handles various configuration files that must be modified whenever a product needs an update (e.g. some of its parameters change, completeness/timeliness expectations vary or new viewers are needed). This, added to the amount of new data flows that EUMETSAT is going to face in the upcoming years, means a considerable workload for the team in charge of its maintenance, which could soon require the availability of more engineers dedicated solely to the configuration and maintenance of SMART.

5.3. SMART required viewers

As of January 2023, there are more than 800 viewers handled by SMART (amongst all the instances), many of them used on a 24/7 basis by ground and spacecraft controllers. This requires a large computational load as well as physical space to display all these viewers.

Each control room has several monitors dedicated to showing a particular set of SMART viewers – with the increasing amount of information displayed, it is becoming rather difficult to perform a visual inspection of all data flows in an effective way.

5.4. Proactive vs Reactive Monitoring

Years ago, when the number of missions and data flows was considerably lower at EUMETSAT, it was possible to base monitoring solely on a proactive basis (that is, performing hourly checks on each of the tools to validate their status). Partly due to the previous challenges above, EUMETSAT increasingly bases its monitoring on a more reactive approach (acting only when alarms are generated).

In fact, SMART can be integrated with GEMS so that it generates alarms when certain thresholds are violated. In this way, the ground/spacecraft controllers do not need to actively monitoring the SMART viewers but simply wait for GEMS audible alarms to be generated. These SMART alarms (potentially hundreds), need to be configured manually, which would also place additional workload on the maintenance teams.

Moreover, their configuration is not always trivial: imagine that we have a product that can be processed in a maximum time of 24 hours. SMART will be configured with a 24h threshold (useful for reporting purposes) and therefore will generate an alarm when this time is exceeded. However, at this moment it will be already too late to react since the product will be already declared as “late”. On the other hand, SMART is configured to arrange products under common services (which are easier to monitor) and the rules for triggering alarms depend on the particular service rather than the products inside.

These challenges and limitations mean that, although reactive monitoring is an ideal approach, it has not yet been fully implemented in EUMETSAT.

5.5. Monitoring overhead

Another challenge associated to the usage of SMART is whether it is really necessary to monitor each point of the product chain. Generally, the most relevant points are the first (i.e. product received at EUMETSAT or L0 generated) and the last ones (i.e. product delivered to users or product received by users, through a particular dissemination mechanism). Therefore, how detailed do we want our monitoring to be for the rest of points? Do we really need to know when the product is e.g. received, pre-processed, processed, renamed and delivered? Or just a subset of all these processes?

Even though being able to monitor all dataflow stages is possible, one has to take into account the additional overhead this generates vs. the actual benefit to the organisation (in general, faster anomaly troubleshooting and resolution). Ideally, each subsystem should make an estimate of how frequently an anomaly in their system can cause an overall dataflow rupture, and how fast should be the resolution of the issue based on what it has been agreed with the end users. In general, the monitoring of new data points should be carefully evaluated to avoid unnecessary overhead.

6. The real need for SMART

In EUMETSAT, monitoring functions can be classified into two large groups:

- **On-event System Monitoring:** the monitoring of the current and past state of all hardware, operating systems, COTS and application level software that form part of the EUMETSAT operational and validation ground segment (GS). This type of monitoring aims at ensuring that every **system** is behaving as expected.
  GEMS and OP5 would belong to this group.

- **Data-flow Monitoring (or End-to-end Service Monitoring):** the monitoring of data as it moves between facilities (or systems), from acquisition to final dissemination/archival. This type of monitoring aims at ensuring that every **service** is nominal, generally in terms of completeness, timeliness and quality.
  SMART would belong to this group.

EUMETSAT now faces a great challenge in terms of monitoring and reporting, largely due to the number of new missions to come. The first question that should be considered before attempting to optimise the tool is: Is the use of SMART as a dataflow monitoring tool **really necessary**?

In many cases, a problem in a particular dataflow (for example, a certain MSG product is received but not disseminated) is detected by both the on-event system monitoring (GEMS generating an alarm in the dissemination facility) and dataflow monitoring (SMART showing a red cell in the viewer indicating a missing product in the dissemination facility). However, that is not always the case; SMART can indicate a real anomaly that goes unnoticed by GEMS in several scenarios:

- A product arriving with a slightly higher timeliness than agreed with the end-users.
- Hanging or slow processes that do not raise alarms but make products arrive later than expected.
- An excessive turnaround timeliness (i.e. time that a product needs to move from one facility to the other within EUMETSAT)
- Number of received products not reaching the minimum required/expected (e.g. EUMETSAT expects at least 20 products per day and we only receive 19).

Based on how the data is handled between GEMS and SMART, the answer to the initial question is clear: **SMART is necessary,** since there are certain unique functions that are not carried out by any other tool at EUMETSAT. However, if we reformulate the initial question in a more abstract way, such as: is the existence of

GEMS and SMART really necessary or can their functionalities be replaced by other (more efficient) tools? The answer is then different.

Since SMART bases its input data on GEMS events, it is possible that if GEMS did some pre-processing before storing events in the database (e.g. could read the event text and classify it already into multiple parameters), SMART would no longer be necessary. This is, in fact, one of the challenges that EUMETSAT is currently facing: trying to reshape the way monitoring and reporting is performed by considering the required functions from scratch.

There are certain COTS (some of them already used in EUMETSAT for some mission-specific data processing monitoring) which are strong candidates for complementing or even substituting GEMS and SMART in the future. They offer integrated search, collection and analysis of large amounts of data, as well as the possibility of applying Machine Learning algorithms to evaluate trends and detect anomalies autonomously. This advanced data analysis component is very attractive, since the post-processing of data coming from SMART is mostly done offline (during the creation of KPI reports) rather than in real-time.

EUMETSAT is currently in the process of determining whether it is feasible to incorporate a new COTS solution into its monitoring and reporting baseline. This requires a detailed study of the current requirements and an extensive analysis of the tools available on the market, in order to verify whether there is one that really fits the organisation. Nonetheless, the use of Machine Learning for the analysis of large volumes of data can already be applied to SMART and GEMS. Even though they do not contemplate this technology in their initial design, it is possible to create ad-hoc tools that allow further post-processing of their data, expanding its current capabilities.

Whatever the future of these tools, the application of ML to real operational scenarios presents large potential benefits. The following sections focus on explaining the status of Machine Learning in EUMETSAT and how the application of this technology, in a practical SMART use case, could indeed be very beneficial for the organisation.

7. Machine Learning at EUMETSAT

Currently, EUMETSAT is in the process of expanding the applicability and utility of Artificial Intelligence (AI) and Machine Learning (ML) to the upstream part of the weather/climate value chain (i.e. thinning of satellite data, gap analysis, retrieval of L2 data) to strengthen the products. However, other departments and tools (such as GEMS and SMART) can also benefit from the use of such technologies.

Some of the current objectives of the organisation in terms of machine learning are to:

- **Foster collaboration:** cooperation between machine learning experts and the scientific community is key to keep up to date with new technologies. Also understanding the way other satellite agencies use ML and creating direct communication channels focused on this topic.
- **Improve data and software infrastructure:** To make the most of the use of ML applied to EUMETSAT, the organisation needs to establish a supporting infrastructure (e.g. cloud service, enough storage resources, etc.) to be able to carry out tests, retrieval and post-processing of data.
- **Support ML training:** The provision of training at all levels, from managers to scientists/engineers, on how this technology can be applied will allow user cases to be created in multiple departments. This will enable problem solving using ML and further foster this technology in the organisation.

It is expected that in the 2023-2025 frame, EUMETSAT will be able to incorporate ML and AI as a basis for satellite data processing and become a reference and collaborator for other space organisations.

8. Machine Learning applied to EUMETSAT operations

Although currently the application of ML in EUMETSAT is mostly centred on satellite data processing, there is also room for its application in operations. The remainder of this paper reviews a use case focused on the configuration of a new dataflow monitoring system, which could also benefit from the use of ML.

8.1. Problem definition

As stated in section 5.2, SMART configuration is becoming very laborious in view of new missions or simply products that become obsolete. As of January 2023, one of the most time-consuming tasks of the EUMETSAT Multi-Mission Performance Analyst is the configuration of SMART caused by the addition/removal of products.

We can take as an example the monitoring of ADM-Aeolus wind profiles (see figure 7).

The Aeolus L2b BUFR files are pushed by ECMWF to the EUMETSAT Multi-Mission Dissemination System (MMDS). Then the data is provided by MMDS for pickup by the EUMETSAT Common Processing Facility (CPF), where it is repacked and pushed back to MMDS. Finally, these products are forwarded to EUMETCast and GTS for dissemination to the end users.

As each product travels through the different facilities, events (or logs) are generated in the system and sent to the GEMS server hosted by MASIF. Each of these events provides information on whether the product has been received, processed or sent to a specific server.

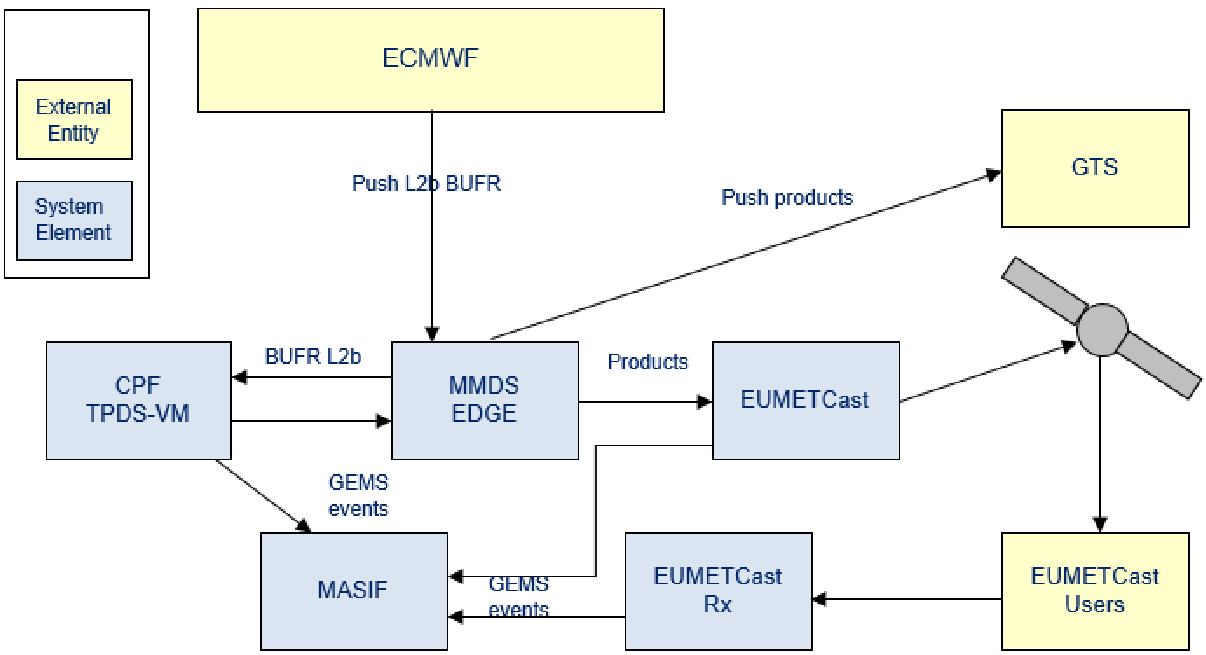

Fig. 7. ADM-Aeolus Retransmission Service Diagram

8.2. Dataset definition

When the ADM-Aeolus dataflow is added into the EUMETSAT ground segment, each server generates a series of GEMS events representing a specific point in the chain. The table below summarizes each of these trackers (or data points), together with the GEMS facility where it is generated and GEMS message associated.

Tracker Name in SMART	**GEMS Facility**	**GEMS event text (simplified example)**
Product received at MMDS from ECMW	MME_EEDGE_OPE	AEOLUS: Aeolus Products: Received file *AE_OPER_ALD_B_N_2B_20230103T120000_20230103T130000_0001.BUFR* at: 23.003.12.38.58 GMT+00:00, size: 8040379 bytes.

Product polled by CPF	MME_CPF_OPE_AEOLUS	AEOLUS: Incoming Aeolus File from MMDS: Polled file *AE_OPER_ALD_B_N_2B_20230103T120000_20230103T130000_0001.BUFR* at: 23.003.12.39.58 GMT+00:00, size: 4906997 bytes, from MMDS-AEOLUS to CPF-OPE-TPDS in 1 sec, 0 retries, 118 sec turnaround time.
Product pushed from CPF to MMDS	MME_CPF_OPE_AEOLUS	AEOLUS: Outgoing Aeolus Data: Sent file *W_XX-ECMWF-Reading,SOUNDING+SATELLITE,ADM+ALADIN_C_ECMF_20230103120000_0001.bin* at: 23.003.12.40.58 GMT+00:00, size: 4909437 bytes, from CPF-OPE-TPDS to MMDS-GTS in 1 sec, 0 retries, 28 sec turnaround time.
Product sent from MMDS to GTS	MME_EEDGE_OPE	GTS_OUT: GTS Data to RMDCN Prime: Sent file *W_XX-ECMWF-Reading,SOUNDING+SATELLITE,ADM+ALADIN_C_ECMF_20230103120000_0001.bin.bz2* at: 23.003.12.41.58 GMT+00:00, size: 4123695 bytes, from MMDS-GTS to GTS-OUT-ALL in 4 sec, 0 retries, 14 sec turnaround time.
Product sent from MMDS to ECast Satellite	MME_ECAST_OPE_UPL	Entry detected:VRB:2023-01-03 12:53:56.999:Delivered file *W_XX-ECMWF-Reading,SOUNDING+SATELLITE,ADM+ALADIN_C_ECMF_20230103120000_0001.bin* id 5ec38gt701918603 from channel 'E1H-TPL-1'
Product received at ECast Reference Station	MME_ECAST_OPE_DL	Entry detected:VRB:2020-01-03 12:53:57.181:Delivered file *W_XX-ECMWF-Reading,SOUNDING+SATELLITE,ADM+ALADIN_C_ECMF_20230103120000_0001.bin* id 5ec38gt701918603 from channel 'E1H-TPL-1'

Table 1 - Trackers and associated GEMS events for the ADM-Aeolus retransmission system

As it can be observed in table 1, six trackers can be configured in SMART from four different GEMS facilities. Each GEMS event includes the *filename* of the ADM-Aeolus product, which contain information about the satellite, product level, sensing start time, sensing stop time, format, etc.

In general, data is well-structured (six types of GEMS events, using two different filename conventions) and follows a constant pattern for each ADM-Aeolus circulated through EUMETSAT.

8.3. Expected ML outputs and challenges

The ML application could be configured to perform different tasks:

1. **Timeliness comparison**. For instance:

   a) Generating automatic alarms when the product timeliness is exceeded. The timeliness can be obtained by comparing the GEMS event timestamp with the product start time (extracted from the filename). While this functionality is not very useful in this particular use case (Aeolus products could be late without requiring the immediate action of a ground controller), it might be interesting in some other missions. Certain EUMETSAT products are directly linked to others (e.g. the processing of *X* does not start until *Y* is generated) and therefore an immediate reaction would be highly desired not to break multiple data chains.
   b) Generating automatic alarms when the product timeliness is exceeded a certain number of times over a period. Since Aeolus products are just redistributed by EUMETSAT, our responsibility falls mostly on the last phases (archival and dissemination to users). Therefore, it would be interesting to know when the products are received late for a continuous period (for example, 50% of the daily cycles) before raising any alarm and triggering internal investigations.
   c) Generating aggregation reports of turnaround timeliness (i.e. time between different facilities) or product timeliness (i.e. event time minus sensing time) over a specific period, useful for KPI reporting.

2. **Detection of dataflow fluctuations**. For instance:

   a) Generating automatic product expectations after detecting a new dataflow across multiple GEMS facilities. This generation could be done autonomously (based on historical data and the number of products received over a period) or with the aid of external tools (such as a product database). In the latter case, EUMETSAT teams could add these expectations (number of products, timeliness, filename structure, etc.) in a basic database so that SMART or the ML application can take into account and recognise new data flows easily.

b) Generating alarms when a new product is found in GEMS, so the engineers can add it to SMART.
c) Generating alarms when a product is found missing in the entire chain, so that engineers can remove it from SMART (when necessary).
d) Generating alarms when a product is found missing from a certain GEMS facility, to trigger anomaly investigations.

Not all the tasks described above encompass the same difficulty. For instance, task *1a* is probably not as complex as task *2a*, which requires a certain volume of training data before the ML application can detect and create an accurate product expectation. In practice, one would start with basic ML outputs that make engineers' work easier, rather than trying to fully automate a very complex task from the very beginning.

In addition, there is an aspect that makes task *2* particularly challenging: the engineering required for converting text based log information to product transit based numerical values. The process of transforming text into numbers is also referred as Natural Language Processing (NLP), which is a branch of AI aimed at giving computers the ability to understand text. The fact that task *2* requires NLP adds an additional layer of complexity, compared to ML simply applied to satellite data processing (already using numerical values as input data).

Another relevant challenge is the data used to train the prototype ML model. The detection of a new data flow might seem incompatible with the use of a large volume of training data, since this flow has just started and there are few GEMS events that can support it. On the other hand, it is generally at the beginning (when the data flow started) where a more extensive monitoring is required, in order to ensure that all the teams have set up their systems accordingly. While it is possible to manually modify GEMS events so that they can be used as test data, it loses part of the automation desired in such an application.

A limitation associated to the automatic detection of data flows is also the fact that initial services are not always predictable from the satellite perspective (especially during LEOP and commissioning phases). For instance, spacecraft often change the mode of operation, activate / deactivate instruments and perform sporadic tests to check certain components. This poses an extra difficulty when providing monitoring and reporting at such a point in the mission.

EUMETSAT has started a prototype application using ML that covers some of the points mentioned in this section. Due to the small volume of training data used and the limited ML expertise that exists in the organisation (only a small group of individuals may be able to solve such challenges), the application is still in development. Nonetheless, in the last years EUMETSAT has participated in ML workshops with other organisation (such as CNES), to share knowledge and use cases, as well as to make other members increase their knowledge in this field.

Following the EUMETSAT ML roadmap expected in the next 3-5 years, it is expected that this ML prototype will soon be able to bring real benefits to the operations teams.

9. Collaboration with other agencies for future monitoring and reporting solutions

While tools such as GEMS and SMART have been tailored to EUMETSAT's needs for mission performance monitoring and reporting, other organisations and external partners probably share similar requirements. One of the objectives of the organisation in the coming years is to strengthen the relationship with space organisations such as ESA, DLR, CNES or NOAA/NASA to create synergies in this topic. The main objective will be to learn how monitoring and reporting is being performed in other agencies in order to re-evaluate existing requirements, learn about other COTS and proprietary solutions, and benefit from the knowledge of external ML experts.

Although EUMETSAT is unique in many ways, there are tasks that are common to many organisations (such as the handling of large volume of data or logs). For instance, there are many products that EUMETSAT receives from external organisations and are simply redistributed internally. Therefore, these products are probably also monitored by these organisations using tools similar to GEMS and SMART.

Summarising, it is fundamental that any consideration of a new tool or platform for mission performance monitoring and reporting shall take into consideration key external partners and organisations (as well as internal stakeholders), to enhance knowledge sharing and the gathering of new ideas.

10. Conclusions

This paper firstly presented the significant amount of new missions and data volume handled by EUMETSAT, which is expected to increase exponentially in the coming years. This is a sign that EUMETSAT continues to grow and offer new services to users, but it also creates challenges that must be resolved.

GEMS and SMART are currently central tools in the monitoring and reporting of satellite products and critical ground segment components, despite presenting several limitations. One of major challenges for SMART is the number of changes that are required for it to work smoothly in operations; all these configuration changes, either to add new missions/products or to maintain the current ones, already occupy practically 100% of an engineer's time.

This paper initially discusses a fundamental question: Is dedicated dataflow monitoring (i.e. the usage of SMART) really necessary? In the current EUMETSAT context, the answer is clear: Yes. SMART and dataflow service monitoring in general is an essential function and therefore cannot be simply removed.

Then, a second question is also considered: Could dataflow service monitoring and on-event system monitoring be combined or simplified in any way? Since these two types of monitoring share certain similarities, it is possible that in the future a single tool (either GEMS and SMART combined, or a different proprietary or COTS solution) could perform both tasks simultaneously. The decision can only be made after EUMETSAT finalises a detailed study of the current requirements and an extensive analysis of the tools available on the market.

Due to all the limitations and challenges currently affecting SMART and GEMS, EUMETSAT is focused on two primary missions in terms of monitoring and reporting: Firstly, try to extend the life of GEMS and SMART in the most optimal way possible, for instance by using ad-hoc ML applications to post-process their output data and complement some of their functionalities. As it has been discussed in this paper, some complex monitoring tasks (such as automatic generation of product expectations) could indeed be covered with the use of this technology.

Secondly, to perform a re-evaluation of the current requirements and existing tools on the market, in order to determine if there is any simpler and more affordable tool (or overall system) that can also carry out the monitoring and reporting functions required by the organisation.

References

[1] T. Edwards, Dealing with the Big Data - The Challenges for Modern Mission Monitoring and Reporting, AIAA 2018-2507, 2018 SpaceOps Conference, Marseille, France, 2018, 28 May - 1 June.

[2] EUMETSAT Operational Services Specification, version 1C, https://www.eumetsat.int/media/45745.

[3] Report on AI/ML Workshop Outcomes, EUMETSAT Internal document, EUM/STG-SWG/52/22/VWG/14, 17 February 2022.

[4] What is natural language processing? - https://www.ibm.com/topics/natural-language-processing (last accessed 06/01/2023)