SpaceOps-2025, ID # 196

DataX: Pioneering data strategies for enhanced mission operations

Evridiki V. Ntagiou[a]*, James Eggleston[a], Pete Collins[a], Katarzyna Cichecka[a]

[a] *Department of Engineering and Innovation, European Space Operations Centre (ESOC) / ESA, Germany*
* Corresponding Author, evridiki.ntagiou@esa.int

Abstract

The European Space Operations Centre (ESOC) oversees a range of complex missions, generating large volumes of diverse data managed by decentralized systems. While effective for mission operations, this decentralized approach poses challenges for data integration, access control, and scalability, limiting the growth of advanced analytics and AI applications. To overcome these challenges, ESOC is developing a unified data layer that treats data and analytics as critical assets within a space mission operations environment. A comprehensive analysis was conducted to evaluate various data architectures and cloud-native solutions. The proposed design supports multiple data types and formats, providing tools for data ingestion, processing, storage, and analysis. Additionally, robust data governance practices are being implemented to ensure data quality, consistency, and security. This paper will present the findings of this initiative, describing how the new data architecture meets the unique demands of a mission operations centre and outlining future steps to enhance and scale the system for evolving data needs.

Keywords: Data architecture, data governance, ground segment engineering, mission operations

Acronyms/Abbreviations

AI	Artificial Intelligence
ACID	Atomicity, Consistency, Isolation, and Durability
BI	Business Intelligence
ESOC	European Space Operations Centre
LLM	Large Language Model
MBSE	Model Based Systems Engineering
ML	Machine Learning

1. Introduction

The European Space Operations Centre (ESOC) currently operates within a fragmented data architecture that, while effective for managing complex operations, presents significant challenges for users such as scientists, engineers, and Artificial Intelligence (AI) systems. These challenges include difficulties with data integration, limited access control, and scalability issues, which hinder the implementation of automated solutions at the system level. The decentralized nature of the system makes it difficult to scale advanced analytics and AI applications, as data is not easily accessible in a usable format, as it may have been specialised to mission specific needs. Furthermore, the continued reliance on legacy systems complicates the handling of the large and diverse datasets generated by modern space missions.

To overcome these obstacles, ESOC is adopting a more scalable approach by treating data and analytics as valuable, reusable assets – a practice known as data assetization. Central to this initiative is the development of a common data layer to simplify data management and integration [1], facilitating the implementation of AI and automation capabilities across ground segment engineering, mission operations and space safety domains [4] [5] [6]. In designing this new architecture, ESOC has evaluated various data models and cloud-native capabilities, focusing on three key areas:

1. **Value-Driven Approach**: Collaborating with data users to understand their needs and use cases – such as system monitoring and reporting—to ensure the common data layer delivers practical value.
2. **Architecture and Tooling**: Designing a flexible architecture that accommodates diverse data types and formats, supported by tools for data ingestion, processing, storage, and analytics. This approach balances the use of cloud-specific services with cloud-agnostic, open-source solutions.
3. **Data Governance**: Implementing a governance framework that ensures data quality, consistency, and security through well-defined rules, policies, and procedures.

The result is a value-driven common data layer and governance model tailored to ESOC's specific needs. Future efforts will focus on refining the architecture and expanding its capabilities to meet the organization's evolving data and operational requirements.

2. Methodology and Deep Dives

ESOC aims to enable scalable data and analytics use cases to enhance automation and aid the goal of improving efficiency within space mission operations. This requires a fundamental shift from managing data in isolated silos to developing scalable data and analytics products. The objective is to establish a more effective and integrated data infrastructure that meets the needs of various operational functions.

A major challenge in this transition is the recurring issue of "reinventing the wheel." Each mission often develops its own set of specialized tools, despite many of these tools serving similar purposes. This duplication leads to inefficiencies, wasted resources, and a lack of consistency across missions.

To address these challenges, ESOC is implementing a common *data layer*: a unified framework designed to build modular, reusable assets that accelerate new developments and analyses. Traditionally, operational data is confined within siloed systems, making it difficult to access and use across different functions. The common data layer breaks down these silos, enabling seamless data sharing and more efficient use of existing tools and analytics. This approach enhances efficiency, fosters consistency, and supports scalable development and deployment of data and analytics products while minimizing redundant efforts.

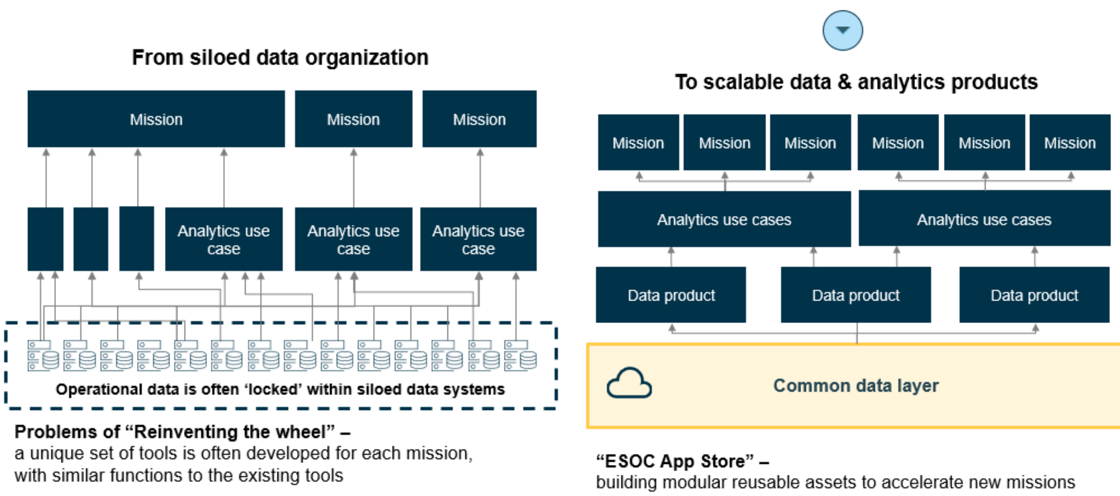


Figure 1. Data organisation vision: from silos to scalable data products

2.1 Pain Points

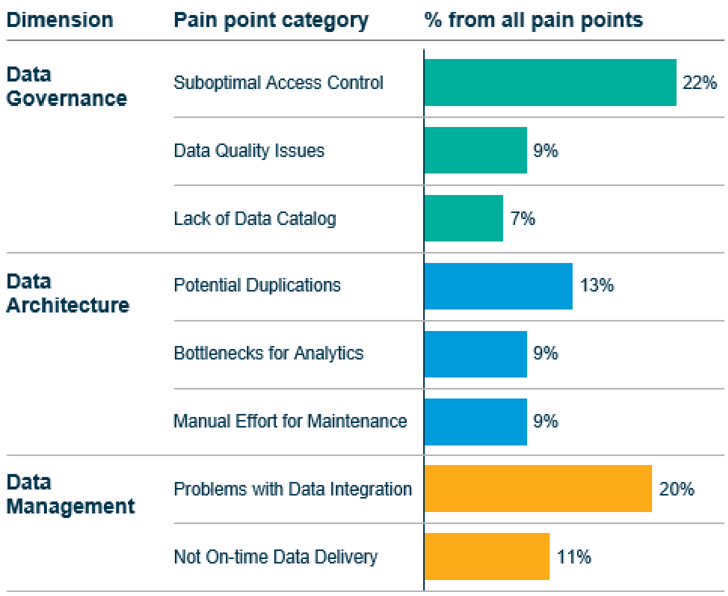
To ensure a comprehensive understanding of the challenges, a detailed analysis was conducted by gathering critical pain points that hinder key data-related activities. This involved one-on-one interviews with personnel across multiple mission operations domains, including:

- Missions (25%)
- Ground Stations and Infrastructure (58%)
- Space Safety (12%)
- Product Assurance (5%)

The most significant pain points identified centred around limited data access and challenges in integrating multiple data sources. The primary cause of these issues was traced to the fragmented system landscape – a collection of diverse systems with overlapping functionalities that complicate navigation and usage.

Additionally, the distinct work processes across different missions and domains create silos in how data is generated, processed, and stored. This fragmentation makes it difficult to integrate and automate data workflows efficiently. The lack of a unified governance framework exacerbates these problems, especially as ESOC faces increasing volumes of data, new missions, and diverse project requirements – all while maintaining a fixed staffing level.

These challenges highlight the urgent need to improve efficiency and modernize current workflows. Implementing a common data layer will help address these issues by facilitating better data integration, streamlining processes, and enabling more automated and scalable data handling across ESOC's operations.



Dimension	Pain point category	Description
Data Governance	Suboptimal Access Control	Fragmented processes for access granting (e.g. emails) and login
	Data Quality Issues	Inconsistencies in data, e.g. for the same parameters across data systems
	Lack of Data Catalog	Lack of centralized wiki on available data sources and their owners
Data Architecture	Potential Duplications	Some data sources potentially have overlapping functionalities
	Bottlenecks for Analytics	Architecture doesn't seamlessly facilitate advanced analytics use cases
	Manual Effort for Maintenance	Legacy systems require a lot of manual efforts for maintenance
Data Management	Problems with Data Integration	Switching between different data systems on a daily basis; non-standardized data format
	Not On-time Data Delivery	Users request a more frequent data delivery; additional data requests take longer to complete

Table 1. Pain points classification and description

2.2 User-Centric Perspective on Data Target State

2.2.1 Mission Operations

Input from mission operations representatives revealed that integrating multiple data sources into existing applications is crucial for enhancing productivity. However, one of the main challenges lies in understanding the meaning of the data itself, whilst accepting and supporting the need of missions and users to modify data types to satisfy specific needs. A key priority identified was the development of applications for anomaly detection and automated report generation. External stakeholders also highlighted significant difficulties in collecting and retrieving data, resulting in labour-intensive software development processes. Additional challenges include:

- Limited seamless integration between systems.

- Lack of standardized data documentation.
- Bottlenecks with on-premises infrastructure, which is difficult to scale.

These findings emphasize the need for better data integration and scalable infrastructure. Based on these insights, the following applications have been prioritized:

- Multi-Source Data Storage: A system enabling the storage of data from multiple sources, directly integrated with analytics toolkits. This would foster an experimental environment for ad-hoc analytics across various data sets.
- Data Catalogue with Data Lineage and Quality Mechanisms: A comprehensive overview of available data systems, including metadata and data ownership. This catalogue could be integrated into dashboards to facilitate data navigation.
- Enhanced Domain Search: Implementing advanced algorithms to search across domain-specific documentation and diverse data sources, improving data retrieval efficiency.
- Query and Reference Documentation Tool: A conversational interface for querying documentation. Solutions using large language models (LLMs) may require a scalable environment to deliver better results.
- Standardized Report Generation and Business Intelligence Tool: An automated tool for generating weekly or monthly reports on spacecraft and ground segment operations. This would integrate multiple data sources to provide a comprehensive overview and generate insights automatically.
- Dashboard for Anomaly Detection and Root-Cause Investigation: A user-friendly application combining multiple data sources, such as telemetry, procedures, and event logs to enhance anomaly detection and root-cause analysis.

2.2.2 Ground Stations

Discussions with ground station operators and engineers identified data system monitoring and maintenance as time-consuming yet essential for ensuring system functionality and data accuracy. A major challenge is the complexity of connecting fragmented systems, where even a proof of concept (PoC) can take up to 12 months to develop.
Based on these insights, the following applications were prioritized:

- Central Data Storage: A centralized storage system with standardized access management tools to store large-scale data (e.g., telemetry, weather information, and time-series logs). This system would improve data accessibility, security, and management.
- Systems Monitoring: Implementing open-source monitoring tools, such as Prometheus and Grafana, to enable multi-dimensional data collection, querying, and alerting. Participants expressed interest in expanding these tools for real-time system monitoring and maintenance visualization.
- Cross-Referencing and Data Access: Addressing the need for bidirectional access between mission data and ground station operations (e.g., retrieving satellite command intentions). Currently, this process is manual and time-intensive. A system enabling automated cross-referencing would significantly improve efficiency.

These applications address the core operational challenges and support the future scalability of ground station activities.

2.2.3 Flight Dynamics

Engagements with flight dynamics stakeholders emphasized the need for more efficient data management, greater automation across processes, and reducing the time required for data preparation and algorithm development. Stakeholders also highlighted future opportunities in Model-Based System Engineering (MBSE).
The following applications were prioritized:

- Data Catalogue with Advanced Search: A centralized data catalogue equipped with advanced search functionality. This tool would improve navigation across scattered information sources (e.g., emails, SharePoint, OneDrive) and prioritize spacecraft-related documentation.
- Model-Based System Engineering (MBSE): Transitioning from document-based to model-based system engineering, allowing for structured and digitally processable representations (e.g., diagrams and tables). This approach enhances automation, reduces time spent on data preparation, and serves as a single source of truth to prevent outdated information. For example, MBSE can automate processes like document generation for signal acquisition and ground station coverage analysis following a satellite separation event. This shift will improve data consistency and efficiency across flight dynamics operations.

These prioritized applications aim to streamline flight dynamics processes, enhance automation, and enable future innovation.

2.2.4 Broader Stakeholder Insights

Additional engagements with Space Safety teams and external ESOC data users revealed similar needs for:

- Improved data accessibility across functions.
- Streamlined data governance to manage increasing data volumes.
- Enhanced tools for data analysis and automation across diverse operational domains.

These insights reinforce the urgency of establishing a common data layer to address cross-domain challenges and support future growth in space operations.

3. Defining a Target Data Architecture

When designing the target data architecture, two critical decisions must be addressed, how to store and query data and how consolidated and unified the data should be. Each decision carries specific implications and requires careful evaluation based on organizational needs and long-term goals.

3.1 Key Value Levers

Developing a modern data infrastructure for space Mission Operations Centres relies on several key value levers:

1. <u>Data Integration and Interoperability</u>
   Seamless integration of data from diverse sources, such as satellites, sensors, and ground systems, enables comprehensive situational awareness and real-time decision-making. Ensuring interoperability across platforms fosters better collaboration and operational efficiency.

2. <u>Real-Time Data Processing and Analytics</u>
   The ability to process and analyze large volumes of data in real time is essential for mission-critical operations. This enables faster anomaly detection, automated responses, and predictive analytics, improving mission performance and safety.

3. <u>Scalability and Flexibility</u>
   With ESOC's mission volume expected to double within the next decade, alongside the rise of industry-led mega-constellations, scalable data solutions are vital. A flexible infrastructure allows for growth, supports new technologies, and accommodates diverse data types without requiring significant system overhauls.

4. <u>Data and AI Application Qualification and Security</u>
   Due to the sensitive nature of mission data, robust cybersecurity measures are crucial [2]. Ensuring compliance with international regulations protects data integrity and enhances operational safety [3].

5. <u>High Availability and Redundancy</u>
   Reliable infrastructure with built-in redundancy is critical to ensure continuous mission operations. This prevents data loss and ensures operational continuity during system failures or disruptions.

6. <u>Automation and Efficiency</u>
   Automating routine tasks—such as data ingestion, cataloging, and validation—reduces human error and allows personnel to focus on higher-value tasks like mission analysis and decision-making.

7. <u>Cost Optimization</u>
   Implementing cost-effective solutions (such as cloud-based or hybrid models) reduces capital expenditures while providing the flexibility to scale resources according to mission demands.

By leveraging these value drivers, mission operations centers can establish a resilient and adaptable data architecture that supports real-time analytics, operational continuity, and future growth.

3.2 How to Store and Query Data

Choosing the right data storage and querying solution is a fundamental architectural decision. Three primary approaches are available:

1. Data Warehouse
    - Description: A centralized system optimized for storing structured data and supporting business intelligence (BI) and reporting.
    - Strengths: Ideal for well-defined data models and structured/semi-structured data.
    - Weaknesses: poor flexibility when faced with heterogeneous or new data types
    - Use Case: Best suited for mission control systems with clear data needs.

2. Data Lake
    - Description: A flexible storage system capable of handling structured, semi-structured, and unstructured data from diverse sources.
    - Strengths: Supports a broad range AI/ML workloads and advanced analytics.
    - Weaknesses: high data quality will require a strong governance approach, system performance cannot be optimised,
    - Use Case: Ideal for data science applications and managing diverse datasets.

3. Data Lakehouse
    - Description: A hybrid approach combining the strengths of data warehouses and data lakes. It provides a unified platform for handling varied data types while supporting analytics and machine learning.
    - Strengths: Offers Atomicity, Consistency, Isolation, and Durability (ACID) transactions, real-time capabilities, and integrated governance
    - Weaknesses: relatively new concept, may not integrate well with existing systems and workflows
    - Use Case: Optimal for handling large-scale mission data while enabling real-time insights and scalable analytics.

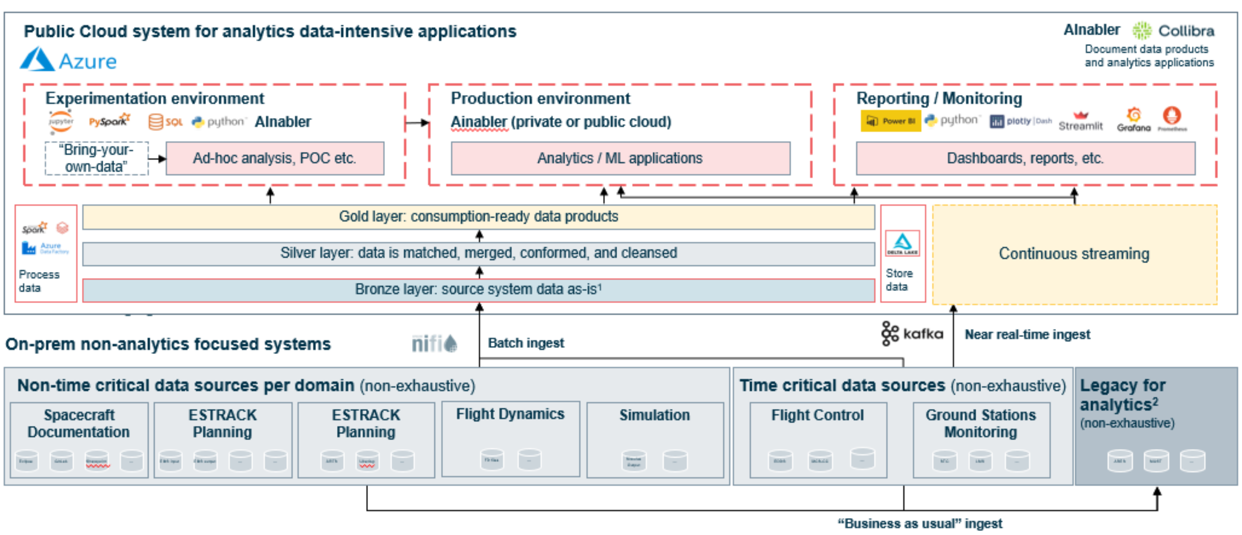


Figure 2. High level target data architecture

3.3 How Consolidated and Unified Should the Data Be?

The degree of data consolidation is another key decision. Two primary architectural paradigms are:

1. Data Mesh
    - *Description*: A decentralized approach where business domains independently manage their own data products using standardized interfaces.
    - *Governance Models*:
        - Fully Federated: Domains operate autonomously, fostering rapid innovation but posing challenges in enforcing global standards.
        - Fully Governed: Central governance ensures standardization across domains, but may reduce agility.
        - Hybrid Federated: Balances autonomy with central oversight, allowing flexibility while maintaining organizational standards.
    - *Use Case*: Suitable for large, mature organizations requiring domain-specific data ownership.

2. Data Fabric
    - *Description*: An integrated approach offering centralized data management across diverse systems, facilitating consistent data discovery, access, and processing.
    - *Topologies*:
        - Data Virtualization: Provides real-time access without moving data but may face performance and complexity challenges.
        - Common Data Repository: Ensures data integrity and security but increases storage costs and requires extensive management.
    - *Use Case*: Ideal for organizations needing centralized control and data consistency across platforms.

3.4 Technology Stack

When selecting a technology stack, considerations include avoiding vendor lock-in and maintaining compatibility with industry standards. A balanced approach is recommended:

- Cloud-First Strategy: Prioritize cloud storage for scalability.
- Best-of-Breed Tools: Use specialized tools tailored to specific needs.
- Exit Strategy: Ensure a clear migration path to prevent long-term vendor dependency.

This approach offers the following benefits:

- Solution Sharing: Enables collaboration with the European space industry.
- Reduced Vendor Lock-In: Minimizes reliance on a single cloud provider.
- Simplified Architecture: Facilitates the integration of specialized tools.
- Comprehensive Support: Leverages cloud provider resources for performance optimization and security.

Compute and storage strategies can be managed independently to ensure adaptability across different platforms.

3.5 Proposed Data Architecture for ESOC and Space Mission Operations Centres

A hybrid cloud with a lakehouse architecture and a hybrid data mesh is proposed as the ideal solution for ESOC's data needs. This model combines:

- Data Lakehouse: Provides scalability, real-time analytics, and ACID transactions.
- Hybrid Cloud: Ensures sensitive data remains on-premises while supporting cloud bursting for scalable workloads.
- Hybrid Federated Mesh: Empowers mature domains to manage their data independently while maintaining central oversight.

The architecture adopts a medallion pattern to organize data into three layers:

- Bronze Layer: Raw data storage.
- Silver Layer: Cleaned and refined data.
- Gold Layer: Curated, high-quality data for analysis and reporting.

3.6 Execution Plan

A detailed execution plan has been developed to guide the implementation of the target architecture. This plan outlines:

- Timeline and Milestones: Key deliverables and project phases.
- Tooling Considerations: Selection of the most suitable technologies.
- Resource Requirements: Computational and human resource planning.

Future developments at ESOC will follow this roadmap, ensuring a structured and scalable rollout.

3.7 Change Management

Successful data transformation requires effective change management to address technical, procedural, and human factors. This involves:

- Impact Assessment: Evaluating how the new architecture affects current workflows.
- Stakeholder Alignment: Ensuring all teams are informed and aligned with the transformation goals.
- Training Programs: Providing personnel with the skills to operate new systems.

A structured change management approach will drive adoption, minimize disruption, and ensure long-term success in implementing the target data architecture.

4. Discussion and Summary

The European Space Operations Centre (ESOC) manages a diverse set of complex missions, generating vast amounts of data through decentralized systems. While this approach is effective for mission operations, it poses challenges in data integration, access control, and scalability, which in turn hinder the growth of advanced analytics and AI applications.

To address these issues, ESOC is developing a unified data layer that treats data and analytics as strategic assets for space mission operations. A comprehensive analysis was conducted to evaluate various data architectures and cloud-native solutions, leading to a new design tailored to handle different data types and formats. The architecture incorporates tools for data ingestion, processing, storage, and analysis, along with governance mechanisms to ensure data quality, consistency, and security.

This paper has outlined the methodology and findings of this initiative, demonstrating how a data lakehouse architecture, combined with a medallion data model, aligns with the specific needs of mission operations. Additionally, it highlights the next steps for refining and expanding the system to support future data requirements.

4.1 Democratization of Mission Operations Data

This initiative will facilitate the democratization of mission operations data, enabling transparent access and structured mapping of data products for both ESA internal stakeholders and external partners. By enhancing accessibility, this approach fosters rapid innovation and strengthens the competitiveness of Europe's Space AI ecosystem [7].

Key benefits include:

- Scalable AI Development: Supports the creation of AI-driven solutions across multiple missions.
- Multi-Mission Capabilities: Encourages system reuse and interoperability between different space missions.
- Increased Efficiency through Automation: Reduces manual workload, allowing mission teams to manage more missions and handle a broader range of tasks effectively.

4.2 Industrialization of the Data Flow

The industrialization of the data flow focuses on standardizing and automating data processes across missions to facilitate efficiency, scalability, and consistency.
By streamlining data ingestion, processing, and management, this effort:

- Reduces Manual Interventions: Minimizes human errors and enhances operational efficiency.
- Enables Seamless Data Integration: Facilitates the integration of data from multiple sources.
- Supports Large-Scale AI Product Development: Provides a structured foundation for expanding AI-driven solutions across ESA's mission portfolio.
- Optimizes Mission Operations: Improves resource utilization and productivity across the European space sector.

This approach will ultimately enhance ESOC's ability to handle an increasing number of missions, making data-driven decision-making a core part of future space operations.

References

[1] E. V. Ntagiou, J. Eggleston, P. Collins, H. Dreihahn, F. Renk, K. Cichecka, DataX: a state-of-the-art data strategy for mission operations, IAC-24-B6,2, 75th International Astronautical Congress, Milan, Italy, 2024, 14 – 18 October.
[2] K. Kotowski, P. Wilczyński, D. Płudowski, A. Kaczmarek, R. Shendy, J. Nalepa, P. Biecek, E. V. Ntagiou, Towards Explainable and Secure AI for Space Mission Operations, SpaceOps ID # 150, 18th International Conference on Space Operations, Montreal, Canada, 26 - 30 May 2025.
[3] ESA, "Machine Learning Handbook," ESA, ECSS-E-HB-40-02A, 2024. [Online]. Available: https://ecss.nl/home/ecss-e-hb-40-02a-15-november-2024/.
[4] G. De Canio, E. Ntagiou, F. Antonello, J. Eggleston, Artificial Intelligence for mission operations automation roadmap: the European Space Operations Centre updates and vision, SpaceOps ID # 372, 18th International Conference on Space Operations, Montreal, Canada, 26 - 30 May 2025.
[5] E. Ntagiou, T. Nogueira, M. Baab, L. Palladino, P. Leitão, L. Medina, B. Werner, C. Texeira, J. Eggleston, Smart Space Operations: OCAI's Contribution to Operational Excellence, IAF Space Operations Symposium, Milan, Italy, 14-18 October 2024.
[6] G. De Canio, A. Riise, S Mejri, V. D. Tran, I. Tanco, R. Semola, V. De Caro, Advancing satellite health monitoring and control with AI, SpaceOps ID # 334, 18th International Conference on Space Operations, Montreal, Canada, 26 - 30 May 2025.
[7] S. Madhogaria, M. Nidugala, E. Cocci, B. Werner, M. Taheran, E. V. Ntagiou, Bridging ESA Innovations with Commercial Space Operations: Integration of an Ops Companion for Satellite Decision Support, SpaceOps ID # 378, 18th International Conference on Space Operations, Montreal, Canada, 26 - 30 May 2025