DATACOM INDUSTRY ASSOCIATION

Emerging Aerospace-enabled Services

The Edge-Cloud Continuum Beyond the Atmosphere

Aerospace WG

© 2026 Datacom Industry Association (DIA).
All rights reserved. Reproduction or distribution, in whole or in part, is permitted with attribution.

Executive Summary

The dominant design philosophy of mobile and satellite networks has long been connectivity-centric: maximise throughput, minimise latency, extend coverage. This paradigm served well through successive generations of cellular infrastructure. Yet the service landscape of the 2020s exposes its structural inadequacy. Emerging applications — autonomous aerial systems, real-time Earth observation analytics, immersive telepresence, distributed AI inference — do not primarily demand higher bitrates. They demand coordinated, adaptive fulfilment of composite service outcomes that span communication links, compute nodes, and cached data assets simultaneously.

Current network architectures respond to this challenge with ad hoc integration: *Multi-Access Edge Computing* (MEC) overlays computation on top of a communication substrate; *Content Delivery Networks* (CDNs) bolt on caching at the application layer; AI pipelines are inserted as separate orchestration planes. The result is a fragmented stack in which the three primary network resources — *Communication, Computing, and Caching* (3C) — are optimised in isolation, without a shared cognitive model of service intent or a unified control logic that spans resource types and network tiers.

The deployment of 3C networks, integrating device, edge, cloud, and communications into a ubiquitous mesh, is moving beyond terrestrial boundaries. By equipping *Low Earth Orbit* (LEO) satellites, *High-Altitude Platform Stations* (HAPS) and aircraft, with communication modems, compute power, and storage, the aerospace ecosystem is evolving into a fully integrated edge-cloud continuum spanning from the stratosphere to low Earth orbit.

This whitepaper, produced by the DIA Aerospace Working Group, identifies emerging data services that may be deployed in Aerospace platforms powered by novel network paradigms in which Communication, Computing, and Caching resources are jointly managed as a unified cognitive substrate.

We define aerospace as the convergence of aeronautics, near-space, and astronautics: a spectrum spanning from tropospheric drones and commercial aircraft to satellites operating beyond the 100 km Kármán line. In the near-space realm, HAPS such as the Airbus Zephyr occupy a uniquely persistent position in the stratosphere at altitudes of 18–30 km.

While this work aligns with the *Non-Terrestrial Network* (NTN) vision championed by the European Space Agency (ESA) NTN Forum and NetworldEurope European Technology Platform, the DIA Aerospace WG takes a distinct architectural stance. Rather than integrating aerospace assets directly into 5G/6G standards per 3GPP Releases 17 and 18, we treat these platforms as standard Internet devices — enabling seamless extension of edge-to-cloud services into the sky, independent of the underlying access technology.

This paper characterizes three high-value service categories - Internet Services, Earth Observation (EO), and Cyber-Physical Services - in terms of their potential addressable markets, governance considerations, interoperability requirements, sustainability implications, and technology challenges.

Table of Contents

1. **Introduction** 5
2. **Aerospace-enabled Services** 6
   - 2.1. Internet Services 6
   - 2.2. Earth Observation Services 8
   - 2.3. Cyber-Physical Services 8
   - 2.4. Summary 9
3. **Impact Factor** 11
   - 3.1. Addressable Markets 11
   - 3.2. Integration Challenges 11
   - 3.3. Policy and Governance 13
   - 3.4. Interoperability 13
   - 3.5. Sustainability 14
   - 3.6. Resilience and Failure Modes 15
   - 3.7. Summary 16
4. **Technological Design Choices** 18
   - 4.1. Architectural Considerations 18
   - 4.2. Protocol Considerations 20
   - 4.3. Orchestration Considerations 24
   - 4.4. Summary 25
5. **Conclusion** 28

List of Contributors

Name	Affiliation
Paulo Mendes	Airbus
Rute Sofia	Fortiss GmbH
Edison Pignaton de Freitas	Halmstad University
Nitinder Mohan	Delft University of Technology
Dirk Trossen	DaPaDot Tech UG
Gorry Fairhurst	University of Aberdeen

1. Introduction

Aerospace infrastructure has long underpinned global communications, but largely in a passive relay role. Today, a confluence of technological advances is enabling a fundamentally different vision: one in which satellites, high-altitude platforms, and aircraft can have an active role not only in forwarding data, but also in storing, processing, and orchestrating it.

The emergence of mega-constellations such as Starlink (SpaceX) and Kuiper (Amazon), the operationalization of HAPS platforms by Airbus (Zephyr/AALTO) and SoftBank (HAPSMobile), and the proliferation of 5G-capable avionics are converging to create a new three-dimensional computing fabric. This ***Aerospace Edge-Cloud Continuum*** extends the terrestrial computing stack vertically into the stratosphere and orbital space.

From a datacom perspective, any flying platform can participate in this continuum if it carries three foundational pillars: Communication modems with multi-link support (GEO/MEO/LEO satellite, Link 16, LDACS, 5G NTN); On-board computing with certified, containerised software execution environments; and On-board storage using radiation-tolerant solid-state media. These pillars transform aerospace assets from passive relays into active Internet nodes.

The DIA Aerospace WG perspective treats aerospace assets as standard Internet endpoints rather than specialized 3GPP NTN nodes. This architectural stance enables Edge-Cloud-Hybrid Cloud service models, such as microservices, serverless functions, and AI inference pipelines, to be deployed and orchestrated across the full edge-cloud continuum using the same abstractions as terrestrial edge computing. The implications for service delivery, data sovereignty, and operational resilience are profound.

This whitepaper explores three service verticals - Internet Services, Earth Observation, and Cyber-Physical Services – that may highly benefit from aerospace datacom integration, while facing some technical, regulatory, and architectural challenges, which must be overcome to realize their potential.

2. Aerospace-enabled Services

This section identifies and characterizes a set of emerging services that may benefit from the integration of aerospace datacom assets in the Internet. Each service is analyzed for technical feasibility, the specific aerospace platform best suited to deliver it, and the unique value proposition relative to purely terrestrial alternatives.

2.1. Internet Services

Aerospace platforms may reshape Internet infrastructure through three primary mechanisms: direct connectivity, asynchronous data transport, and distributed edge computing.

Global Connectivity and 5G Non-Terrestrial Networks

LEO constellations provide global broadband coverage with round-trip latencies of 20 ms to 60 ms, while Medium Earth Orbit (MEO) systems offer an intermediate option at roughly 125 ms to 250 ms, compared to the 600+ ms typical of Geostationary Earth Orbit (GEO) systems. In this context, aircraft and HAPS can serve as relay nodes or direct access providers.

HAPS, operating in the stratosphere between 18 and 30 km, may deliver persistent, high-speed, low-latency coverage over targeted geographic areas without the orbital mechanics constraints of satellites. A single HAPS aircraft such as the Airbus Zephyr can cover approximately 7,500 km² with low-latency direct-to-device connectivity.

When looking at the NTN concept, aircraft equipped with 5G NTN modems (3GPP Release 17/18) can be used to bridge LEO satellite segments and terrestrial users, reducing dependence on expensive ground station infrastructure.

In either case (aircraft, HAPS), aircraft to satellite Free-Space Optical (FSO) links, as well as inter-aircraft optical links, may be used to aggregate traffic collected by 5G NTN modems, increasing the capacity of the NTN system.

Moreover, aerospace nodes can act as “Aerial eXchange Points” similar to Internet eXchange Points enabling interactions and integration of several satellite orbits, ground nodes and aerial nodes through a combination of FSO and radio links.

Data Muling — Asynchronous High-Capacity Transport

A less-discussed but economically compelling service is the physical transport of massive datasets via commercial aircraft. Airlines operate globally, connecting data centers across continents with high frequency and predictable scheduling. By outfitting a commercial fleet with high-capacity data storage, an aircraft fleet may achieve aggregate data rate value that rivals or exceeds terrestrial network bandwidth for bulk transfers.

For instance, a fleet of 5,000 aircraft each carrying 64 TB of certified storage yields a combined airborne capacity of 320 PB, constituting a highly secure, energy-efficient transport system for immense data loads such as genomic databases, satellite imagery archives, and AI model checkpoints.

While unsuitable for real-time workloads, data muling achieves effective throughput orders of magnitude higher than any wireless link for bulk transfers exceeding 10 PB, making

it competitive for synchronising globally distributed data centres.

Aerospace Edge Computing and In-Orbit Data Centres

The sky may become a decentralised data centre. Modern LEO satellites may be equipped with space-qualified GPUs and AI accelerators to run foundation models and perform complex analytics directly in orbit. The emerging *Orbital Edge Computing* (OEC) paradigm may transform LEO constellations into distributed computing platforms capable of hosting serverless workloads, AI inference, and real-time analytics.

The in-orbit data centre market is projected to reach $39.09 billion by 2035, growing at a Compound Annual Growth Rate (CAGR) of 67.4% from $1.77 billion in 2029, driven by AI compute demand, solar energy abundance, and passive radiative cooling that eliminates terrestrial cooling costs.

On the one hand, the EU-funded ASCEND programme (Advanced Space Cloud for European Net zero emissions and Data sovereignty), funded at €300 million through 2027, is developing orbital data centre modules targeting a first in-orbit demonstration in 2028. ASCEND envisions platforms twice the size of the ISS, powered by megawatt-scale solar arrays.

On the other hand, companies including Starcloud (NVIDIA Inception partner), Axiom Space, and Lonestar Data Holdings are racing to deploy commercial computing payloads in LEO. Starcloud plans to launch an orbital data centre with 100× the GPU compute ever previously deployed in space in 2026.

In regards to aircraft, a fleet of aircraft with modern AI inference hardware could collectively provide *Peta-Operations per Second* (POPS) of distributed compute, processing sensor data at the edge to dramatically reduce air-to-ground downlink requirements.

Global-scale In-Orbit Inference Distribution Networks (iIDNs)

Alongside the computational capabilities of in-orbit data centres comes the in-orbit storage and distribution of (AI) inferencing knowledge that can serve large-scale geo areas, covered by in-orbit objects at any point in time. Coupled with contextual capabilities to place the knowledge where most relevant, in-orbit Inference Distribution Networks (IDNs) can frontload but also complement terrestrial IDN capabilities in order to enable large-scale but also geographically limited AI services. Combined with localised fusion and AI capabilities, the inference knowledge can be specialised and adapted with locally available knowledge, further enhancing the value of those iIDNs.

A concrete example regards to Earth Observation Services, as described in the next session, such as wildfire response with no terrestrial connectivity. A computer-vision model trained to detect smoke and fire progression from multispectral imagery is served within the iIDN rather than on the ground. When a satellite or HAPS overflies the area, inference runs in orbit against the locally held model and only the result — a geolocated detection, a confidence score, and a predicted spread vector — is delivered to first responders, instead of downlinking the full image to a terrestrial data centre and waiting for the analysis to return. This collapses the time from observation to actionable alert from minutes to seconds and removes the dependence on a ground link that may be congested or destroyed by the event itself. Because the model is distributed across the in-orbit objects covering the region, coverage persists as individual platforms move out of view, and the model can be specialised in place with geography-specific priors.

2.2. Earth Observation Services

Earth Observation is one of the most mature and immediately deployable aerospace service verticals. The integration of multi-layer platforms — satellites, HAPS, and aircraft — enables a new generation of persistent, multi-spectral, AI-processed observation services that transcend what any single platform can achieve.

Ultra-High Resolution and Persistent Observation

Commercial LEO satellites typically achieve spatial resolutions of 30–50 cm. Platforms like Planet Labs' SkySat constellation deliver sub-50 cm resolution with daily revisit rates globally.

On the one hand, aircraft operating at lower altitudes may achieve resolutions of 5–10 cm, enabling applications such as precision infrastructure inspection, urban mapping at cadastral detail, and high-fidelity crop monitoring.

On the other hand, Airbus' Strat-Observer service, flying on the Zephyr HAPS at 20+ km altitude, delivers 18 cm resolution imagery over 2,500 km² per day and continuous near-real-time video. The larger Zephyr variant, expected in 2026, doubles payload capacity, further enhancing EO capabilities. AALTO achieved a 67-day stratospheric endurance record in April 2025, demonstrating the persistence possible with HAPS.

Multi-Sensor Fusion and Onboard AI

The true power of aerospace EO lies not in individual sensor performance but in the fusion of complementary modalities. Aircraft and HAPS may carry heavier and more diverse payloads than many satellite classes, enabling simultaneous collection of optical, LiDAR, hyperspectral, infrared, and Synthetic Aperture Radar (SAR) data.

For instance, onboard AI processors may run geospatial foundation models in real time, enabling autonomous identification of anomalies — wildfires, flood boundaries, oil spills, or crop stress — without waiting for data downlink.

Data from LEO satellites can be fused with high-fidelity aircraft or HAPS observation data using multi-modal AI to produce insights inaccessible to any single platform.

Processing data onboard and transmitting only derived insights, rather than raw imagery, helps to reduce downlink bandwidth by orders of magnitude, which is an important factor taken into account the constraint communication of LEO satellites.

2.3. Cyber-Physical Services

Aerospace platforms are uniquely positioned to act as low-latency edge nodes for coordinating and managing cyber-physical systems at regional and continental scale. Their persistent, elevated vantage point enables sensing and actuation capabilities unavailable to terrestrial infrastructure.

Smart Mobility and Cooperative Perception

Aircraft and HAPS flying over metropolitan areas can provide continuous wide-area situational awareness to support ground-based *Advanced Driver Assistance Systems* (ADAS) and autonomous vehicle fleets.

In this context, edge AI on aerospace nodes may be used to analyze real-time traffic flow, detect hazards (debris, incidents, congestion), and broadcast contextual awareness to vehicle networks with latencies compatible with safety-critical applications.

This new aerial perspective may be used to complement onboard vehicle sensors, which are limited by occlusion and short range, providing a broader context for route optimization and collision avoidance.

Such a scenario requires the integration of aerospace platforms with *Vehicle-to-Everything* (V2X) communication standards, enabling aerospace nodes to serve as regional traffic orchestrators, dynamically adjusting signal timing and routing in response to detected anomalies.

Decentralized Air Traffic Management

As *Urban Air Mobility* (UAM) and autonomous drone operations scale, centralized *Air Traffic Management* (ATM) becomes a bottleneck. Aerospace datacom platforms can enable a distributed, AI-mediated air traffic management paradigm.

Aircraft and HAPS may serve as relay and compute nodes for a distributed ATM mesh, enabling autonomous negotiation of routes, speeds, and separations between aircraft.

In this scenario, distributed ledgers may provide tamper-resistant, auditable records of airspace allocation and conflict resolution decisions, eliminating single points of failure, while AI agents in each aircraft may apply local optimization metrics to globally communicated constraints, converging on an optimized solution without centralized authority.

Situational Awareness and Critical Infrastructure Monitoring

Persistent HAPS coverage may enable continuous monitoring of critical infrastructure (pipelines, power grids, coastlines) with rapid anomaly detection. Commercial aircraft, equipped with downward-facing sensing and communication antennas, may complement this with recurrent monitoring along their regular flight corridors, providing frequent revisits of infrastructure beneath dense air routes even though, unlike HAPS, they cannot loiter persistently over a fixed area, while multi-domain sensor fusion from satellite, HAPS, and aircraft may enable correlation of events across spatial scales, from regional weather patterns to localized structural deformation.

When monitoring critical infrastructures, secure, encrypted communications channels between aerospace nodes and ground operators may support defence and civil protection missions under denied or congested communication conditions.

2.4. Summary

The three service verticals are not equally mature. Internet Services, anchored by LEO connectivity and data muling, are the most immediately deployable, with commercial operations already underway. Earth Observation is the most technically coherent vertical — the combination of persistent HAPS coverage, onboard AI, and multi-sensor fusion ad-

dresses real market gaps that neither satellites nor aircraft alone can fill. Cyber-Physical Services carry the highest long-term value but also the steepest integration complexity, particularly where safety-critical latency requirements constrain which aerospace platforms can meaningfully participate.

A common theme across all three verticals is that the value proposition of aerospace platforms is complementary to terrestrial infrastructure. The case for aerospace integration is strongest where terrestrial alternatives are absent, constrained, or prohibitively expensive, not where they already perform well.

Key Takeaways

- No single platform tier delivers the full service potential of any vertical. Multi-layer integration is the value proposition, not individual platform capability.
- Data muling's competitive framing requires correction: it competes on cost-per-bit for latency-insensitive bulk transfers, not on throughput against live fiber.
- Cyber-physical services with explicit safety boundaries: aerospace nodes augment terrestrial and short-range infrastructure; they do not replace it

KPIs and Research Directions

Service	KPI	Target	Direction
Connectivity	Round-trip latency to end device	<60ms (LEO), <10ms (HAPS/Aircraft)	5G NTN modem integration; FSO internode links
Data Muling	Cost per petabyte transported	Competitive with long-haul fiber for >10PB	Fleet storage certification; Ground ingestion automation
EO Resolution	Onboard AI inference latency	< 5s from collection to derived insight	Geospatial foundation model compression for edge hardware
EO Fusion	Cross-platform data latency	< 30min satellite-to-HAPS/aircraft-to-ground fusion cycle	Standardised EO data interchange formats
Cyber Physical	HAPS / Aircraft cooperative perception latency	< 50ms end-to-end under operational conditions	Deterministic 5G NTN scheduling; onboard inference HAPS/aircraft, FSO backhaul as fallback

3. Impact Factor

This section analyses the potential impact of the identified service verticals on value-added datacom markets, examining addressable market scale, integration complexity, policy constraints, interoperability requirements, and sustainability implications.

3.1. Addressable Markets

The financial potential of the aerospace 3C continuum spans multiple high-value verticals. The following table summarizes key market estimates for the primary service domains identified in Section 2.

Service Domain	Market Size (2025)	Projected Size	CAGR
Global IoT Connectivity	~$630B (overall IoT)	$1.5T–2.65T by 2030	~18–20%
Aircraft IoT Services	~$11B	$78B by 2033	~23%
HAPS Platforms	$6.4B	$11.76B by 2032	~9%
In-Orbit Data Centres	$1.77B (2029 baseline)	$39.09B by 2035	67.4%
Earth Observation (HAPS/Aircraft)	~$5B	$13.3B by 2032	~12%
Smart Mobility / V2X	~$45B	$124B by 2030	~18–22%

Table 1: Addressable Market Overview for Aerospace-Enabled Services (2025–2035)

Several dynamics are amplifying these headline figures. The AI computer crisis, with data centre power demand projected to grow 165% by 2030 (Goldman Sachs), is driving urgent interest in orbital data centres as a means to access unlimited solar power and passive cooling.

Defense budgets are sustaining HAPS procurement as persistent *Intelligence, Surveillance, and Reconnaissance* (ISR) platforms, while commercial adoption accelerates for broadband delivery and EO.

For instance, SoftBank's planned commercial HAPS service launch in Japan in 2026 and the NTT Docomo/Space Compass investment of $100M in AALTO represent concrete inflection points for commercial market development.

3.2. Integration Challenges

Physical and operational constraints present some engineering hurdles across all service domains. These challenges span environmental extremes, platform-specific constraints,

and the fundamental complexity of managing highly dynamic, intermittent network topologies.

Environmental and Structural Constraints

In regards to stratospheric operating conditions, HAPS must endure temperatures of −65°C, high solar radiation, and stratospheric wind jets reaching 100+ km/h. The Airbus Zephyr's 25 m wingspan and sub-75 kg weight reflect the engineering extremity required for endurance flight.

In what concerns aircraft, integrating external antennas (e.g., massive MIMO 5G arrays or free-space optical laser turrets) may introduce aerodynamic drag, structural fatigue risks, and *Size, Weight, and Power (SWaP)* constraints. A typical 5G macro base station antenna unit consumes over 11 kW and weighs several hundred kilograms, which may be incompatible with current aircraft electrical and structural limits without fundamental redesign.

In-orbit data centres face extreme thermal cycling (facing the sun vs. eclipse) and must dissipate computational heat without convective cooling, which constitutes a fundamental departure from terrestrial data centre thermal engineering.

Compute and Storage Certification

In the aviation domain, on-board computing systems must satisfy DO-178C (software) and DO-254 (hardware) certification requirements, imposing rigorous verification and validation overhead on software-defined or AI-driven systems not originally designed for aviation.

In space, qualified computing hardware (radiation-tolerant processors, ECC memory, radiation-hardened storage) is significantly more expensive and less performant than Commercial Off-The-Shelf (COTS) equivalents, constraining the compute density achievable per unit mass and cost.

Network Topology Dynamics

In space, LEO satellites at 550 km altitude complete an orbit every ~90 minutes at ~27,000 km/h. Task scheduling must account for satellite visibility windows, inter-satellite link topology changes, and the handover of active workloads between compute nodes.

In the aircraft domain, flight routes are partially predictable but subject to diversions. Workloads initiated on aircraft compute nodes may need seamless migration when an aircraft descends out of the predicted route.

Those topology changes impact not just the distribution algorithms usually realised through routing capabilities but also upper layer capabilities. For instance, topological changes may impact the validity and relevance of stored metadata or AI inferencing data if it becomes irrelevant for the new coverage area, thus requiring linking topology knowledge to the knowledge distribution mechanisms themselves. This, however, also opens up the opportunity to utilise the predictability of topology changes to "pre-load" upper layer knowledge so as to make this knowledge available at the time of entering the new coverage of in-orbit objects.

3.3. Policy and Governance

The aerospace edge-cloud continuum operates at the intersection of aviation regulation, spectrum governance, data protection law, and export control regimes. Navigating this multi-layered governance landscape is a non-trivial prerequisite for commercial deployment.

The *European Union Aviation Safety Agency* (EASA) enforces any modification to a commercial aircraft (including antenna installation, power system changes, or avionics integration) to get an EASA (or FAA) supplemental type certification, a process lasting 2–5 years. EASA is expected to finalize HAPS operational rules in 2026, which will significantly impact European commercial deployment timelines.

In regards to spectrum, HAPS-specific allocations exist at 28, 31, 47–48 GHz, and 2.4/3.5 GHz bands under International Telecommunication Union (ITU) Radio Regulations. The 47 GHz FCC allocation in the US underpins domestic HAPS broadband services. International harmonization of the HAPS spectrum is ongoing within ITU-R Study Group 5.

When talking about data sovereignty, processing personal data in orbit or on aircraft transiting multiple jurisdictions may raise complex General Data Protection Regulation (GDPR) and equivalent questions. Determining applicable law for data processed at 20 km altitude over international waters has no settled legal precedent.

In what concerns privacy and surveillance, providing 5–10 cm resolution persistent imagery from HAPS and aircraft may raise significant privacy concerns requiring robust legal frameworks, consent mechanisms, and anti-censorship protocols for responsible deployment.

In terms of dual-use and export controls, aerospace datacom systems with imaging or communication capabilities may be subject to EU Dual-Use Regulation, ITAR (US), or equivalent national export control regimes, constraining technology transfer and international collaboration.

3.4. Interoperability

For the aerospace edge-cloud continuum to function as a coherent 3C network, its constituent platforms must interoperate seamlessly with terrestrial Internet infrastructure and with each other across heterogeneous link technologies and administrative domains.

In this context a service-centric networking model, in opposition to the current Internet host-to-host packet delivery, must support service discovery, load balancing, and media distribution across dynamically changing network topologies. Protocols such as Routing on Service Addresses, Media over QUIC and other publish-subscriber protocols for decentralised communications, as well as AI agent schema enabling cross-platform workflow composition may be required.

In an NTN context, aerospace nodes may operate as dynamic extensions of terrestrial public and private networks, with seamless handover between satellite, HAPS, aircraft, and ground segments as topologies evolve.

Moreover, for services requiring tight integration with 5G/6G core networks, 3GPP Release 17 (NTN) and Release 18 (NTN enhancements) provide standardized interfaces.

Commercial deployment may require federation of services across operator boundaries,

in a similar way to Internet *Border Gateway Protocol* (BGP) peering or roaming in mobile networks. Standardized service catalogues, trust frameworks, and billing interfaces are prerequisites.

3.5. Sustainability

Aerospace edge computing offers genuine environmental advantages over terrestrial data centres, but these benefits are conditional rather than inherent. Realizing them requires honest lifecycle accounting, responsible orbital stewardship, and a willingness to quantify costs that are easy to omit from a platform-centric analysis.

Operational Advantages

Space-based platforms have continuous access to solar irradiance unaffected by cloud cover, night-time cycles, or seasonal variation, making megawatt-scale solar arrays a credible primary power source. HAPS platforms such as the Airbus Zephyr are solar-electric, producing zero direct CO_2 emissions during flight. The ESA ASCEND programme extends this logic to orbital scale, envisioning platforms powered entirely by solar arrays as a structurally carbon-neutral alternative to fossil-fuelled terrestrial data centre capacity — provided the operational phase is not considered in isolation.

Thermal management offers a parallel advantage. In the vacuum of space, computational heat is dissipated through radiation rather than water cooling or mechanical refrigeration, eliminating a cost that consumes 30% to 40% of power in conventional data centres and imposes substantial freshwater demands. For Earth observation and sensor fusion workloads, onboard processing yields a further gain: transmitting derived insights rather than raw imagery reduces downlink bandwidth requirements by orders of magnitude, lowering both network energy consumption and ground infrastructure load.

Launch Emissions and Lifecycle Accounting

These operational benefits must be weighed against the emissions embodied in reaching orbit. A Falcon 9 mission emits approximately 425 tonnes of CO_2-equivalent while delivering around 22,000 kg to LEO, which corresponds to roughly 19 kg CO_2 per kilogram of payload. A modest 300-satellite constellation with five-year operational lifetimes requires approximately 60 replenishment launches per decade, producing around 25,500 tonnes of CO_2 in launch emissions alone, before manufacturing is considered. Larger constellations scale this figure proportionally.

A less-quantified but potentially significant concern is black carbon deposition. Rocket exhaust injects soot directly into the upper stratosphere, where its radiative forcing effect per unit mass is considerably greater than surface-level emissions. Current atmospheric models characterize this impact poorly, and it is absent from most constellation sustainability assessments.

The consequence is straightforward: aerospace edge computing offers a credible sustainability advantage only when the operational carbon savings — avoided terrestrial data centre energy consumption, reduced downlink bandwidth, eliminated cooling infrastructure — exceed the amortised launch and manufacturing emissions over the full platform lifetime. This break-even calculation has not been rigorously established in the published

literature. The DIA Aerospace WG identifies it as a research priority, and cautions against characterising any orbital platform as carbon-neutral on the basis of operational phase analysis alone. The ESA ASCEND programme's carbon-neutral framing, for instance, requires this fuller accounting before the claim can be substantiated.

Orbital Debris and Long-Term Sustainability

Debris risk has moved from a background concern to an active constraint on LEO deployment. The catalogued debris population currently exceeds 35,000 objects larger than 10 cm, with far greater numbers of smaller, untracked fragments. Congestion in the 550–600 km altitude band — the dominant range for commercial broadband constellations — is approaching the critical density thresholds modelled in Kessler syndrome analyses, beyond which collision-generated debris could trigger a self-sustaining cascade.

The Inter-Agency Space Debris Coordination Committee (IADC) guidelines recommend active deorbit within five years of end-of-life, yet this remains far from universal industry practice. Responsible large-scale LEO deployment requires deorbit capability to be treated as a non-negotiable design requirement rather than an optional mitigation, with its mass, cost, and fuel budget explicitly accounted for in constellation economics. Conjunction analysis, automatic manoeuvre coordination between operators, and adherence to evolving ITU and national debris mitigation frameworks are parallel obligations.

The DIA Aerospace WG recommends that any sustainability assessment of aerospace edge infrastructure adopt a lifecycle framework encompassing launch emissions, manufacturing, operational energy, and end-of-life disposal. Platforms that satisfy this full accounting represent a genuine environmental contribution; those that do not should not be positioned as inherently green alternatives to terrestrial infrastructure.

3.6. Resilience and Failure Modes

The aerospace edge-cloud continuum operates in an environment subject to intermittent connectivity, highly dynamic topologies, hostile environmental conditions, and potentially adversarial threats. Unlike terrestrial cloud infrastructures, aerospace platforms cannot always rely on continuous connectivity, immediate human intervention, or rapid hardware replacement. Consequently, resilience and survivability must be treated as fundamental design principles rather than operational afterthoughts.

Each aerospace platform tier has distinct failure characteristics that the orchestration layer must anticipate:

Platform	Failure Mode	Recovery Timescale	Service Impact
Satellite	Orbital decay, eclipse, handover gap	Orbital cycle (e.g. 90-min LEO)	Workload migration required
HAPS	Stratospheric wind jet displacement, power budget shortfall at night	Hours to days	Coverage gap over target area
Aircraft	Diversion, early descent, ground hold	Minutes to hours	Mid-flight workload orphaning

Platform	Failure Mode	Recovery Timescale	Service Impact
Ground Station	Congestion, physical outage	Variable	Uplink/downlink bottleneck

Key resilience principles worth articulating:

- **Checkpoint-based workload migration.** Stateful compute jobs running on aircraft or HAPS nodes should emit lightweight checkpoints to a persistent store (ground or satellite) at configurable intervals, enabling resumption on an alternative node without full recomputation. Moreover, when an aircraft loses uplink connectivity mid-flight, its onboard computer must continue operating autonomously without creating a conflicting state. *Conflict-Free Replicated Data Types* (CRDTs) or timestamp-ordered logs are candidate mechanisms.
- **Fault-Tolerant Service Continuity**. Service architectures must be designed to tolerate partial system failures without compromising overall mission objectives, which may require: replication of critical services and data across multiple nodes; automatic failover mechanisms capable of redirecting workloads and traffic; distributed state synchronization enabling seamless service continuity during workload migration events; and graceful degradation strategies (e.g. (reduced resolution, increased latency tolerance, local caching only) allowing services to continue operating with reduced functionality, being services classified by criticality — safety-of-life, mission-critical, best-effort.
- **Autonomous Recovery and Self-Healing.** The scale and geographical distribution of the continuum make manual operational intervention impractical. Hence self-healing properties as required to enable autonomous detection, diagnosis, and mitigation of failures, such as: continuous monitoring of infrastructure health, service performance, and resource utilization; AI-assisted anomaly detection capable of identifying emerging failures before service disruption occurs; Automated reconfiguration of communication, compute, and storage resources in response to detected faults; Predictive maintenance mechanisms using telemetry and operational history to anticipate hardware degradation.
- **Cyber Resilience.** As aerospace nodes become active participants in critical digital infrastructure, they may become attractive targets for cyberattacks. Security mechanisms must therefore extend beyond prevention to include the ability to maintain operations during active attacks or component compromise: Isolation of compromised components through dynamic segmentation and containment mechanisms; Continuous verification of software integrity and runtime behavior; Distributed trust frameworks reducing dependence on centralized security authorities; Cryptographic agility enabling the adoption of post-quantum security mechanisms as future threats emerge.

3.7. Summary

The market figures in Section 3.1 are indicative rather than aerospace-specific. The continuum's addressable share of those markets depends on resolving the integration, governance, and sustainability challenges identified in subsequent subsections — none of which are peripheral concerns. Regulatory certification timelines alone - 2 to 5 years per EASA Supplemental Type Certificates (STC) impose a structural lag between technical readiness and commercial deployment that market projections rarely reflect.

The sustainability picture is more complex than the industry narrative suggests. Operational advantages are real but insufficient justification for carbon-neutral claims without full lifecycle accounting. For instance, orbital debris is transitioning from a future risk to a present constraint that must be costed into constellation economics today.

The resilience analysis makes explicit what the service descriptions (in Section 2) imply: the continuum is not a reliable substrate by default. Reliability is an engineered property of the orchestration layer, not a characteristic of the platforms themselves.

Key Takeaways

- Market projections should be treated as upper bounds contingent on governance resolution, not baseline forecasts.
- The sustainability advantage of aerospace edge computing is real but conditional; lifecycle break-even analysis is a prerequisite for credible environmental claims.
- Resilience must be designed in from the architecture layer, not retrofitted; failure modes are predictable and should be pre-mitigated.
- The 3GPP interoperability requires explicit resolution: Internet-native and NTN-integrated stances are complementary at different layers, not competing.

KPIs and Research Directions

Domain	KPI	Target	Direction
Certification	EASA STC cycle time for airborne datacom	< 18 months for defined payload classes	Pre-approved modular payload certification framework
Sustainability	Lifecycle CO_2 per petabyte processed in the aerospace ecosystem	Established break-even vs. terrestrial baseline	Standardised aerospace compute lifecycle assessment methodology
Debris	Active deorbit compliance rate	100% within 5 years of EOL	Deorbit-as-a-service; mandatory deorbit budget in constellation licences
Resilience	Workload recovery time on node loss	< 60s for mission-critical; < 300s for best-effort	Checkpoint interval standardisation; successor node predesignation
Federation	Cross-operator service handover success rate	>99.5% for connectivity services	Standardised service catalogue and trust framework (analogous to roaming)
Data Sovereignty	Jurisdictional compliance automation rate	100% without human intervention per hop	Machine-readable usage policy enforcement at protocol layer

4. Technological Design Choices

This section identifies the major technical challenges inherent in realizing the aerospace service continuum, with particular focus on the impact on Internet architecture, protocol design, and service orchestration. The challenges are intentionally framed at a level of abstraction suitable for driving a multi-year research agenda.

4.1. Architectural Considerations

The classical Internet was engineered around assumptions that the aerospace continuum systematically violates: that hosts have stable addresses, that topology changes occur on timescales of hours or days, and that connectivity between any two points is continuous. In the aerospace continuum, a LEO satellite completes an orbital pass in roughly 90 minutes, a HAPS may drift under stratospheric wind loading, and an aircraft may divert without notice. Nodes that are reachable today may be unreachable in minutes, and the compute node best placed to serve a request at one moment may have passed below the horizon by the time a response is due. Realizing high-throughput, low-latency service delivery across this environment requires rethinking four foundational architectural assumptions: the host-centric addressing model, the implicit trust extended to network nodes, the passivity of network elements with respect to intelligence, and the static nature of resource allocation.

From Host-Centric to Service-Centric Networking

The IP addressing model binds service reachability to host identity and location. When hosts move continuously and disconnect unpredictably, this binding becomes a liability. A service-centric networking model decouples what is being requested from where it currently resides, enabling the continuum to satisfy requests from whichever node is best positioned to do so at the moment of request.

The architectural properties this requires are: named data and service objects that carry identity independent of their current physical location, so that a request for a specific EO dataset or inference result can be resolved to any node holding a valid copy; rapid service registration and withdrawal as nodes enter and leave coverage, maintaining a consistent and current view of available capabilities across the continuum; content-aware routing that matches incoming requests to nodes based on data provenance, computational capability, and proximity rather than a fixed destination address; and load distribution across heterogeneous nodes with differing compute, storage, and link capacities, preventing concentration of demand on nodes that happen to be well-connected at a given moment.

Information-Centric Networking (ICN) architectures, and specifically *Named Data Networking* (NDN), provide a research baseline for this paradigm. Their extension to highly mobile, intermittently connected topologies remains an open problem and a priority research area for the continuum.

Particularly for mixed environments, i.e., where terrestrial combine non-terrestrial segments (e.g., for Internet services or AI inference distribution networks), combining host- and service-centric concepts may be desirable. For instance, solutions such as Routing on Service Addresses[1] combine the benefits of in-data-transfer discovery (of services) with

[1] Dirk Trossen, Luis Contreras, Jens Finkhäuser, and Paulo Mendes, "Architecture for Routing on Service Ad-

the topological foundation of an IPv6 network, allowing backwards-compatible deployments in both Terrestrial Networks (TN) and Non-Terrestrial Networks (NTN) segments, while reaping the benefits of supporting higher service dynamics and lower latencies, through avoiding dedicated discovery steps such as done through the existing *Domain Name Service* (DNS).

Zero Trust Architecture

The physical and administrative diversity of the aerospace continuum precludes any model of implicit trust based on network location or operator identity. A satellite operated by one jurisdiction, a HAPS managed by a second, and a ground station under a third administrative domain may all participate in the same service delivery chain. Any node in that chain may be compromised, misconfigured, or subject to conflicting regulatory obligations. Zero Trust principles — in which trust is never assumed and must be continuously earned through verified identity and demonstrated behaviour — are the only viable security foundation for this environment.

The practical architectural requirements are mutual cryptographic authentication for all inter-node communications, establishing verified identity before any data or control plane interaction; fine-grained authorisation evaluated per service request rather than per session, with policy enforcement embedded at every layer of the stack; continuous behavioural monitoring and anomaly detection capable of identifying compromised or misbehaving nodes and isolating them from the service plane without manual intervention; and encryption of data both in transit and at rest, with key management mechanisms designed explicitly for intermittent connectivity. Moreover, standard PKI assumes synchronous access to a certificate authority — an assumption that breaks down for nodes in communications blackouts. Key management under disconnected operation is not a secondary concern but a first-order architectural requirement.

AI-Native Architecture

The continuum cannot be managed by human operators reacting to topology and system changes in real time. The speed, scale, and heterogeneity of the environment demand that intelligence be embedded at every layer, from orbital nodes performing onboard inference to ground orchestrators optimizing multi-domain execution plans. An AI-native architecture treats machine intelligence as a structural component of the network rather than an application running on top of it.

The key requirements are standardized agent schemas and discovery directories enabling AI agents deployed across different aircraft, satellites, and ground systems to find each other, negotiate capabilities, compose multi-step workflows, and execute tasks without requiring pre-established bilateral integrations. Federated inference distributes decision-making across the continuum rather than centralizing it, reducing latency for time-sensitive decisions and eliminating single points of failure in the control plane. Context-aware agents maintain a model of their own position in the continuum with information such as current resource availability, link quality, QoS obligations, and regulatory constraints, and adapt their behaviour accordingly, without requiring instructions from a central authority for every operational decision.

dresses", Internet Draft (draft-trossen-rtgwg-rosa-arch-01), July 2023

The interoperability of AI agents across platforms operated by different organizations requires open, standardized schemas for agent capability advertisement and task composition. The absence of such standards today is a concrete gap that the DIA Aerospace WG identifies as a near-term standardization priority.

Dynamic Adaptability and Resilience

Static resource allocation — assigning a workload to a node at deployment time and assuming it will remain there — is incompatible with a continuum in which node availability is continuously changing. Architectural mechanisms must treat resource allocation as a continuous process rather than a one-time decision.

This requires seamless migration of stateful workloads as node availability changes: when an aircraft begins descent, when a satellite approaches the edge of its coverage window, or when a HAPS loses solar power at dusk, running workloads must be verified and transferred to successor nodes within the available transition window, with service continuity maintained throughout. Adaptive resource allocation must respond to fluctuating compute, storage, and communication availability in real time, redistributing load before degradation becomes visible at the service layer. Graceful degradation policies must be prespecified for every service, defining the reduced operating modes — lower resolution, increased latency tolerance, local caching only — that activate automatically when node availability falls below threshold, and the recovery procedures that restore full service as nodes re-enter coverage.

Resilience in this architecture is not a property of individual nodes but of the orchestration layer's ability to anticipate, detect, and respond to topology changes faster than service SLAs are breached. The failure mode and resilience analysis in Section 3.6 provides the operational grounding for these architectural requirements.

Nevertheless, due to the predictability that often underlies the movements of in-orbit objects, the architectural concepts and protocols mechanisms should exploit *time variance* as a key input, e.g., allowing for scheduled operations across several layers of the systems for, e.g., routing changes, knowledge distribution mechanisms, handover, and others.

4.2. Protocol Considerations

The traditional TCP/IP stack was designed for a relatively static, terrestrial Internet with stable host addresses, predictable topologies, and low-latency connectivity between endpoints. The aerospace continuum violates all three assumptions simultaneously. Nodes move at high velocities, link availability is intermittent and topology-dependent, and the same physical platform may serve radically different workloads — from safety-critical ATM to background telemetry — within the same operational window. Targeted protocol evolution, and in some cases new protocol designs, are required to address these conditions.

Semantic-Aware Protocols

Protocols operating in the aerospace continuum should comprehend the semantics of the data and services they carry, not merely their bit-level representation. This capability enables three practically valuable behaviours. First, intelligent in-network processing, where data is transformed, aggregated, or filtered as it transits through intermediate nodes, re-

ducing unnecessary onward transmission, particularly valuable given the severe downlink constraints of LEO satellites. Second, semantic routing, in which requests are matched to nodes based on the meaning and provenance of the data sought rather than a destination address, enabling content-aware load distribution across heterogeneous nodes. Third, context-aware protocol behaviour that adapts dynamically to the semantic priority of traffic — prioritising a safety-critical airspace conflict alert over background sensor telemetry without requiring application-layer intervention.

In-Network Computing Protocols

For distributed machine learning and AI inference workloads, protocols should enable computation to occur as data flows through the network rather than only at designated endpoints. This requires gradient aggregation at intermediate nodes rather than end-to-end centralisation, enabling federated learning across the continuum without exposing raw data to any single node. It also requires in-network function execution, where serverless functions are invoked at the node where data already resides, minimising unnecessary data movement across constrained links. Underpinning both is delay-tolerant operation: protocols must buffer, prioritise, and forward data as nodes enter and exit coverage windows, without requiring persistent end-to-end connectivity for correct operation.

Privacy-Preserving Data Exchange

As aerospace nodes transit multiple national jurisdictions within a single operational period, data sovereignty compliance cannot rely solely on administrative controls. It must be enforced at the protocol layer. Continuum nodes should exchange model updates or gradients rather than raw training data, applying federated learning principles to minimise data exposure. Where aggregate statistics must be shared, differential privacy mechanisms should bound the information leakable about any individual data point. Encrypted data in transit should carry machine-readable usage policies that intermediate nodes can enforce, ensuring that data processed over international waters or foreign airspace remains compliant with its originating jurisdiction's requirements without human intervention at each hop.

Latency-Bounded Protocol Design for Cyber-Physical Services

Protocol design for cyber-physical workloads must be validated against concrete latency budgets rather than general low-latency objectives. The relevant application requirements differ substantially: V2X safety messages (ETSI ITS-G5 / C-V2X) require end-to-end delivery within 10ms for collision avoidance; cooperative perception sharing tolerates up to 50ms for ADAS augmentation; *Urban Air Mobility* (UAM) separation assurance requires responses within 200ms per EUROCAE ED-269 guidance; and smart traffic signal control requires sub-100ms intervention to be meaningful.

As an example of the need to bound the protocol latency, let us look at the case of the HAPS-mediated V2X path. The vehicle-to-HAPS interface is inherently bidirectional (5G NTN supports both uplink and downlink over the same air interface), so the relevant design question is whether the request must be relayed to the ground for processing or can be served onboard the HAPS. Because the HAPS is itself an active compute node in this con-

tinuum, two paths must be distinguished: a ground-assisted path, in which the HAPS relays the request to a ground node over an FSO backhaul, and a direct onboard path, in which cooperative-perception inference is executed on the HAPS and the response is returned directly to the vehicle without any ground backhaul. The latency of each decomposes as follows:

Segment	Estimated Latency
Vehicle to HAPS uplink (5G NTN, 20 km altitude)	2–5ms
HAPS onboard processing	1–3ms
HAPS to ground backhaul (FSO)	<1ms
Ground processing and response	2–5ms
Downlink to vehicle	2–5ms
Total round-trip — ground-assisted path	**~8–19ms**
Total round-trip — direct onboard path (no ground backhaul)	**~5–13ms**

The two paths place HAPS-mediated services in different feasibility tiers. The ground-assisted path (~8–19ms) is marginally feasible for ADAS cooperative-perception augmentation under favourable conditions, but outside the deterministic envelope required for hard collision-avoidance safety messages. The direct onboard path (~5–13ms) — in which inference is executed on the HAPS and returned to the vehicle over the same bidirectional 5G NTN interface, with no ground backhaul or ground-processing leg — comfortably meets the 50ms cooperative-perception budget and brings the lower end of the range close to the 10ms collision-avoidance threshold. Even so, the residual variability of the wireless air interface means the direct onboard path cannot yet be relied upon as the sole channel for sub-10ms hard safety functions: the final safety layer for sub-10ms V2X should remain with dedicated short-range communication — Dedicated Short-Range Communication (DSRC) or C-V2X sidelink — operating independently of the aerospace infrastructure. Protocol designs that assume aerospace nodes can fully substitute for this short-range safety layer introduce an unacceptable safety dependency; however, the direct onboard path is the appropriate target architecture for latency-bounded cooperative perception, and ground backhaul should be treated as a fallback rather than the default.

The same decomposition applies to commercial aircraft carrying downward-facing 5G NTN antennas and onboard computers, but the figures are lower rather than merely comparable. At a typical cruise altitude of ~10 km — roughly half the HAPS altitude — the two distance-dependent propagation legs (vehicle-to-platform uplink and platform-to-vehicle downlink) each halve to ~1–2.5ms, while the compute-bound legs (onboard processing, ground processing) are unchanged. The per-leg budget for the aircraft case is therefore as follows:

Segment (aircraft, ~10 km cruise altitude)	Estimated Latency
Vehicle to aircraft uplink (5G NTN, ~10 km altitude)	1–2.5ms
Aircraft onboard processing	1–3ms
Aircraft to ground backhaul (FSO)	<1ms
Ground processing and response	2–5ms
Downlink to vehicle	1–2.5ms
Total round-trip — ground-assisted path	**~5–14ms**
Total round-trip — direct onboard path (no ground backhaul)	**~3–8ms**

The direct onboard path for aircraft (~3–8ms) is materially faster than the HAPS equivalent (~5–13ms), and its entire range now falls at or below the 10ms collision-avoidance budget rather than merely approaching it. This is the qualitative shift: for the aircraft case, even the worst-case direct-path estimate (~8ms) clears the hard-safety threshold, which the HAPS path does not. The caveat established above still holds — these are nominal budgets, and the residual jitter of the wireless air interface, scheduling variability, and handover events mean the figures cannot be treated as a deterministic guarantee. The aircraft direct onboard path is therefore a credible target for hard-safety cooperative perception within its coverage corridor, but a dedicated short-range safety layer (DSRC or C-V2X sidelink) should still backstop it until the bound is demonstrated to hold deterministically under operational conditions. The practical distinction relative to HAPS remains geometric rather than protocol-level: a HAPS loiters persistently over a fixed area, whereas an aircraft serves only the corridor it overflies, making aircraft a complementary low-latency node along dense air routes rather than a persistent regional one.

For LEO-mediated services, the one-way propagation delay of approximately 2.5ms at 550 km altitude, compounded by handover gaps between orbital passes, makes sub-10ms round-trip delivery structurally unachievable. LEO participation in smart mobility services is therefore appropriately scoped to non-safety-critical functions: traffic pattern analytics, route optimization, and infrastructure monitoring, where latency tolerances of seconds rather than milliseconds apply.

Protocol specifications for cyber-physical aerospace services should explicitly state the latency tier they target, the application classes they can and cannot serve, and the terrestrial or short-range fallback mechanisms required for safety-of-life functions.

Protocol Virtualization and Testing

The cost and operational risk of testing protocol behaviour in real aerospace deployments makes virtualisation an essential design discipline rather than an optional validation step. Protocol behaviour should be characterised in high-fidelity virtual sandboxes that model realistic orbits/routes, link intermittency, node heterogeneity, and mobility patterns before any hardware integration. Digital twins for the aerospace continuum, replicating the topology dynamics of specific constellation configurations and HAPS coverage geome-

tries, should be treated as first-class development infrastructure, enabling continuous protocol refinement without operational exposure.

4.3. Orchestration Considerations

Managing a multi-dimensional, constantly reconfiguring edge-cloud continuum requires an orchestration layer that operates at a fundamentally different level of abstraction than conventional cloud resource schedulers. Terrestrial orchestration frameworks such as Kubernetes assume that nodes are reachable, that topology is stable on timescales longer than scheduling cycles, and that the cost of moving data between nodes is low relative to compute cost. None of these assumptions hold reliably in the aerospace continuum. The orchestrator must be intent-driven, topology-aware, latency-bounded, and capable of operating across administrative boundaries it does not control.

Intent-Based Service Orchestration

Rather than accepting low-level resource specifications, the continuum orchestrator should translate high-level service intents — expressed as objectives across latency, throughput, geographic constraint, cost, and regulatory compliance — into concrete execution plans dynamically composed from available nodes. The orchestrator needs a stable home, which becomes possible with different aerospace configurations. A HAPS, persistent over a fixed area for weeks, is well suited to run the regional orchestrator and hold control state, with ground and cloud above it and LEO satellites and aircraft joining as workers when in view. The orchestrator consults a live service catalogue reflecting current node availability, link quality, and compute capacity, and invokes an optimisation engine to generate an execution plan satisfying all stated constraints simultaneously. Critically, this plan must be treated as a living document: as node availability changes due to orbital mechanics, aircraft diversions, or link degradation, the orchestrator must revise the execution plan in real time without interrupting service delivery. Policy enforcement — data sovereignty boundaries, export control constraints, privacy requirements — must be integrated into plan generation rather than applied as a post-hoc filter, ensuring that no execution plan is instantiated that violates regulatory constraint.

Latency-Aware Workload Placement

Consistent with the latency budgets established for protocol consideration, the orchestrator must classify workloads by latency tier at ingestion and enforce placement rules accordingly. Safety-critical cyber-physical workloads requiring sub-10ms response must be rejected from HAPS-mediated or LEO-mediated execution paths and directed to an aircraft based system or a terrestrial infrastructure. Workloads tolerating 50–200ms may be placed on HAPS or aircraft nodes under the latency budgets derived above. Latency-insensitive workloads — bulk analytics, model training, archival processing — are appropriate candidates for LEO or orbital edge placement. The orchestrator should expose latency feasibility analysis to service operators at plan-generation time, surfacing explicit warnings when a requested workload cannot meet its SLA from the available node pool, rather than accepting the request and silently degrading performance.

Data Orchestration and Compute-Data Co-location

A fundamental orchestration decision is whether to move compute to data or data to compute. The cost profile of each strategy differs substantially depending on data volume, available link bandwidth, and node compute capacity. Compute migration — moving a containerised processing function to the node where raw data resides — minimises data movement and is preferred when the source node has sufficient available compute. Data streaming to an available compute node is preferable when the source node is compute-constrained, provided link bandwidth and latency permit. For EO workloads in particular, the data gravity of raw satellite imagery typically favours onboard processing: transmitting a derived anomaly detection result consumes orders of magnitude less downlink capacity than transmitting the underlying imagery. The orchestrator should model both strategies for each workload and select based on real-time resource availability, not a fixed policy.

Workload continuity across node transitions requires explicit orchestration support. When an aircraft descends out of coverage, or a satellite passes below the horizon mid-computation, the orchestrator must detect the imminent transition, initiate workload checkpointing or live migration to a successor node, and complete the handover within the available window. For LEO nodes completing an orbital pass in approximately 90 minutes, this window is predictable and schedulable. For aircraft, route deviations require the orchestrator to maintain contingency placement plans that activate on divergence from the filed flight plan.

Multi-Domain Federation and Policy

Commercial deployment requires orchestration across administrative boundaries between satellite operators, aircraft operators, ground network providers, and cloud infrastructure providers. No single operator controls the full continuum. Federation mechanisms must therefore provide standardised service description and discovery interfaces enabling cross-domain service composition without requiring bilateral integration agreements for each operator pair. Automated SLA negotiation and enforcement must operate between federated domains, with machine-readable contracts specifying performance guarantees, data handling obligations, and liability boundaries. All service execution and resource consumption events should be recorded in auditable, tamper-resistant logs suitable for billing reconciliation, regulatory compliance, and post-incident analysis across domain boundaries.

The federation model for the aerospace continuum is functionally analogous to BGP peering in the terrestrial Internet — establishing reachability and trust between autonomous systems without requiring centralised coordination — and to roaming agreements in mobile networks, which enable service continuity across operator boundaries without full infrastructure integration. Standardised service catalogues, trust frameworks, and billing interfaces equivalent to those underpinning terrestrial Internet federation are a prerequisite for the commercial aerospace continuum, and their development should be a near-term standardisation priority.

4.4. Summary

A service-centric model without zero trust creates a discoverable attack surface. Zero trust without AI-native key management breaks under intermittent connectivity. AI-native orchestration without dynamic adaptability produces brittle automation that fails at the moments it is most needed. The architecture must be designed as an integrated whole,

not as independently deployable components.

The protocol and orchestration analyses converge on a shared insight: the aerospace continuum requires that latency, trust, data sovereignty, and workload placement decisions be made at the infrastructure layer, not delegated to applications. Applications operating in this environment cannot be expected to handle the complexity of intermittent connectivity, jurisdictional compliance, and node migration themselves. The infrastructure must absorb this complexity and present a coherent service abstraction above it.

The gap between current terrestrial cloud-native frameworks and what the aerospace continuum requires is significant but not unbridgeable. The building blocks — service centric networking, federated learning, CRDTs, delay-tolerant networking, intent-based orchestration — exist in research form. The missing elements are integration, standardisation, and validation at aerospace-relevant scale and operating conditions.

Key Takeaways

- The four architectural pillars are interdependent; partial implementation degrades the value of each.
- Latency tier classification must be enforced at the orchestration layer, not left to application developers; the HAPS latency budget analysis provides a concrete baseline for this classification.
- Key management under disconnected operation is the most underspecified requirement in current zero trust frameworks applied to aerospace; it requires dedicated protocol work.
- Digital twin environments for the continuum are not a testing convenience — they are a prerequisite for protocol and orchestration validation given the prohibitive cost of at-scale aerospace testing.
- Standardised AI agent schemas and directories are the single highest-leverage near-term standardisation investment: they unlock cross-platform workflow composition across all three service verticals simultaneously.

KPIs and Research Directions

Area	KPI	Target	Direction
Service Centric Networking	Service discovery latency	<500ms from coverage acquisition to service availability	NDN extension for mobile, intermittent topologies
	Content-aware routing overhead vs. IP baseline	<10% additional latency	Lightweight semantic annotation at the network layer
Zero Trust	Authentication latency under disconnected operation	<200ms using pre-provisioned trust anchors	Short-lived certificate pre-issuance; offline OCSP stapling
	Key freshness guarantee during blackout	Defined policy tiers: full / read-only / emergency by time since CA contact	Tiered trust model standardisation
AI-Native	Agent discovery latency across oper-	< 1s for capability adver-	Open agent schema standardisation (priority deliv-

Area	KPI	Target	Direction
	ator boundaries	tisement and matching	erable 2026–2027)
	Federated inference accuracy vs. centralised baseline	>95% parity at defined latency budget	Gradient compression for constrained aerospace links
Dynamic Adaptability	Stateful workload migration time on node loss	<30s for mission-critical workloads	Checkpoint protocol standardisation; successor pre-designation
Protocol Semantic	In-network processing bandwidth reduction	>80% downlink reduction for EO-derived insight vs. raw imagery	Semantic annotation and in-network aggregation protocols
Protocol Delay Tolerance	Data delivery rate < 50% link availability	> 99% eventual delivery for non-real-time workloads	Delay-Tolerant Networking (DTN) profiling for aerospace topologies
Protocol Privacy	Jurisdictional compliance enforcement latency	Zero additional end-to-end latency (policy enforced inline)	Machine-readable usage policy at protocol layer
Orchestration	Execution plan revision latency on topology change	<10s for non-safety-critical workloads	Predictive topology modelling from orbital mechanics
	Cross-domain SLA negotiation time	<5s automated negotiation for standard service classes	Standardised machine-readable SLA templates
Digital Twin	Fidelity vs. live deployment	<15% deviation in latency and throughput metrics	Orbital mechanics and link intermittency modelling for major constellation configurations

5. Conclusion

The aerospace edge-cloud continuum has crossed the threshold from research vision to commercial programme. LEO mega-constellations are operational, HAPS platforms have demonstrated stratospheric endurance beyond 60 days, orbital compute payloads are entering development, and regulatory bodies are finalising operational frameworks. The foundational infrastructure is being built with or without a coherent architectural framework to guide it.

The risk is not that the continuum fails to materialise — it is that it materialises as a fragmented collection of proprietary, incompatible systems that replicate the siloed architecture of terrestrial MEC, CDN, and AI pipeline stacks at vastly greater complexity and cost. Avoiding this outcome requires that the architectural, protocol, and orchestration principles identified in this whitepaper be pursued as active standardisation priorities in parallel with platform deployment, not as a second-generation clean-up exercise.

The provided sustainability and resilience analyses establish two non-negotiable design constraints that must be embedded in commercial programmes now: lifecycle carbon accounting that includes launch emissions, and deorbit capability treated as a mandatory design requirement rather than an optional mitigation. Both are easier to incorporate at the design stage than to retrofit into operational constellations.

Priority Recommendations for 2026–2030

The DIA Aerospace WG identifies five priorities, ordered by leverage and urgency:

1. Open AI Agent Schema and Directories *(Near-term: 2026–2027)* The highest-leverage near-term standardisation investment. Standardisation for agent capability advertisement, discovery and task composition unlock cross-platform workflow automation across all service verticals without requiring bilateral integration agreements.

2. Intent-Based Orchestration for the 3D Continuum *(Medium-term: 2027–2028)* Extending terrestrial cloud-native orchestration frameworks — Kubernetes, Terraform, service mesh — to handle intermittent connectivity, topology dynamics, and multi-domain federation. Latency enforcement and predictive topology are two critical requirements.

3. In-Network Computing Protocols for Intermittent Topologies *(Medium-term: 2027–2029)* Delay-tolerant networking, gradient aggregation at intermediate nodes, and semantic-aware routing adapted to specific topology classes, such as LEO constellations and HAPS coverage geometries.

4. Key Management Under Disconnected Operation *(Near-term: 2026–2027)* The most underspecified gap in current zero trust frameworks applied to aerospace. Short-lived certificate pre-issuance and tiered trust degradation policies are the concrete deliverables.

5. Lifecycle Sustainability Framework *(Near-term: 2026)* A standardised methodology for full lifecycle carbon accounting of aerospace compute platforms, covering launch, manufacturing, operations, and end-of-life disposal. This is a prerequisite for credible sustainability claims and for regulatory frameworks that may impose carbon disclosure obligations on space operators. The orbital debris component — mandatory deorbit budget costing and IADC compliance verification — should be integrated into constellation licensing frameworks as a parallel deliverable.