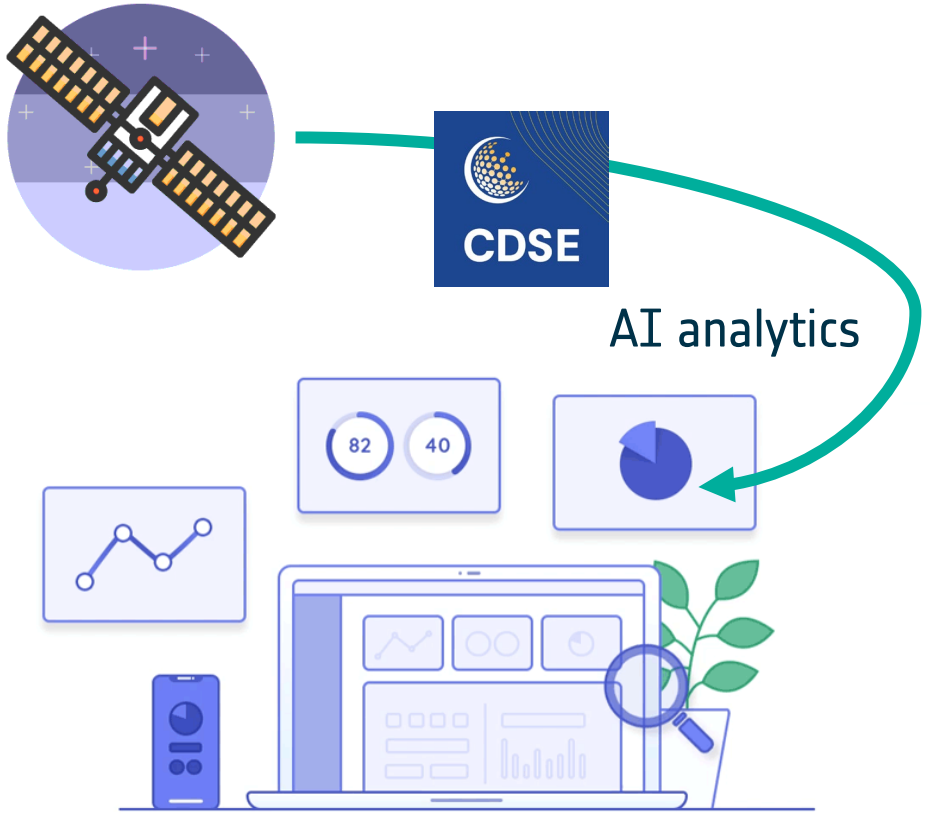
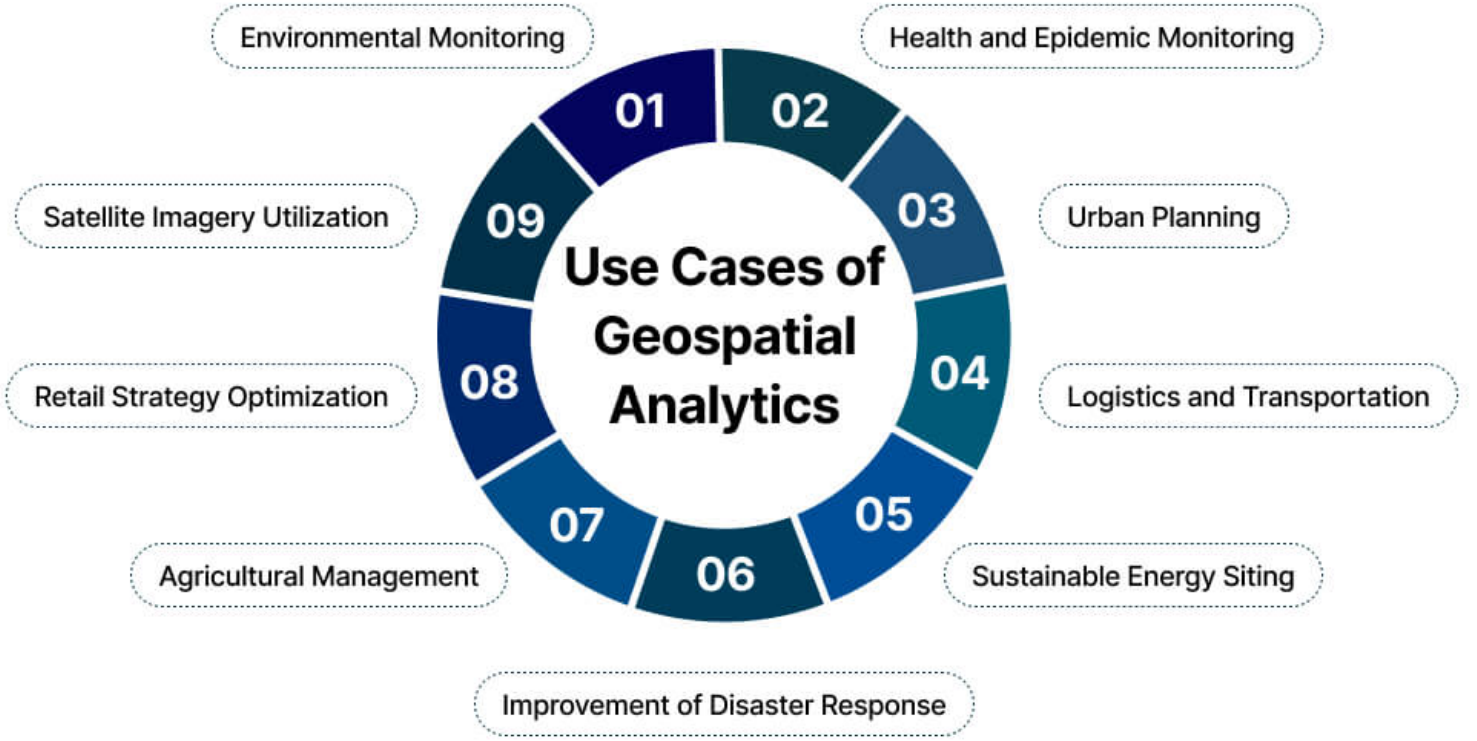
Unlocking the Potential of L1A Sentinel-2 Multispectral Imagery for Onboard Artificial Intelligence: Methodology, Dataset Creation, and Future Prospects

Roberto Del Prete, 7th Sentinel-2 Validation Team Meeting

It's all about the data

Geospatial AI, integrating **artificial intelligence** with spatial data, is rapidly transforming industries through advanced analytics, predictive modeling, and real-time decision-making, driving the market from **$85.77 billion in 2022** to a projected **$226.53** billion by **2030** at a 12.6% CAGR.

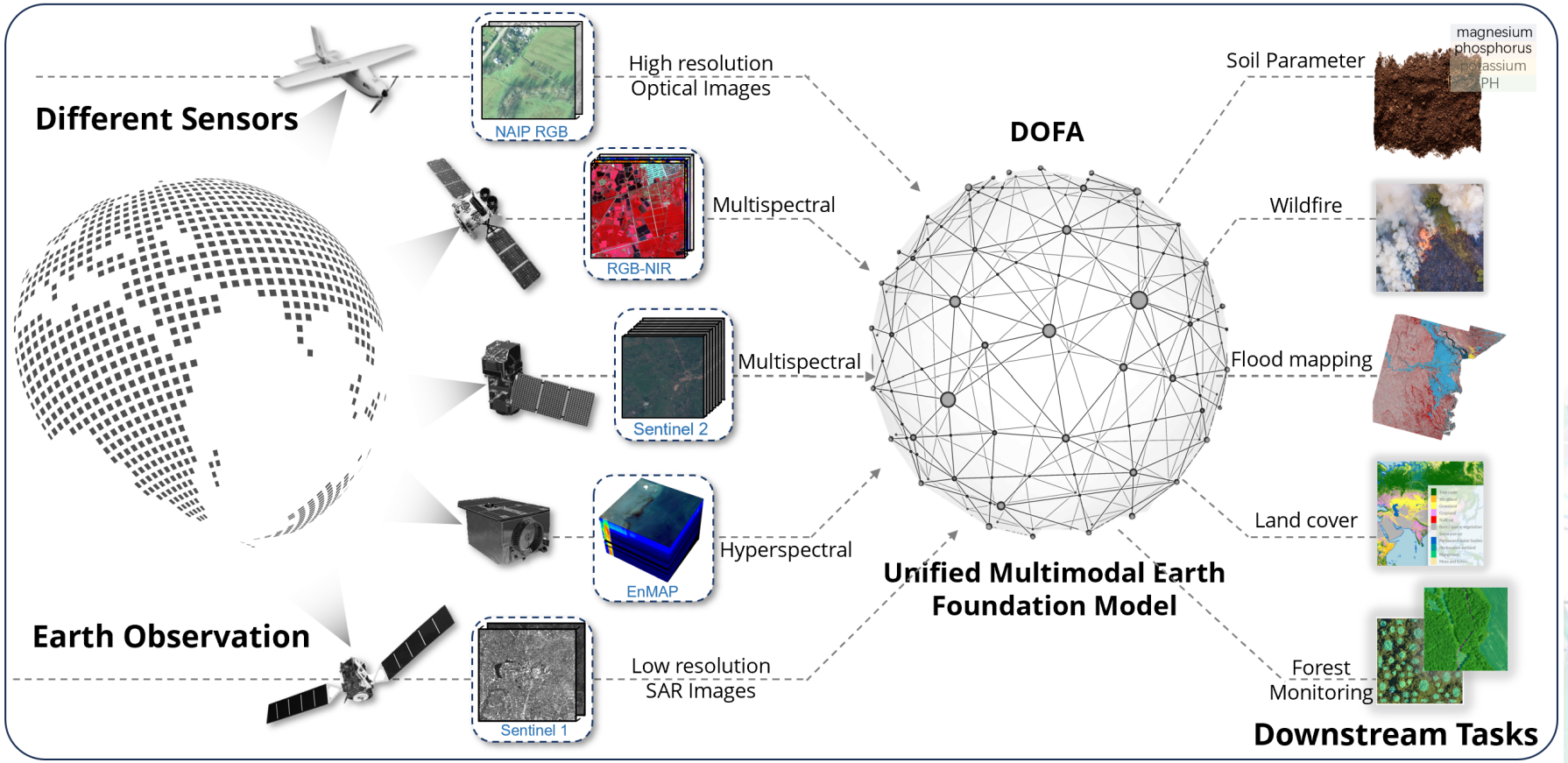




Credits: https://www.solulab.com/geospatial-ai/

The rise of GFMs

In recent times, the exploitation of multiple data sources has proven unprecedented capabilities



By using **SSL**, we are able to train powerful vision encoders directly exploiting data itself.

NO NEED FOR LABELS →

This opened to the **GFMs** aiming for a unified data representation for different downstream tasks.

Key insight:
We are now in an epoch where we can mix information coming from different sources with AI.

To put in simple terms: good models comes from good data (the more the better)

Data-driven AI

- Since 2013, AI and machine learning have seen rapid R&D expansion.
    - AlexNet
- Data volumes and computational power have grown exponentially.
- These advances enabled training of increasingly powerful models.

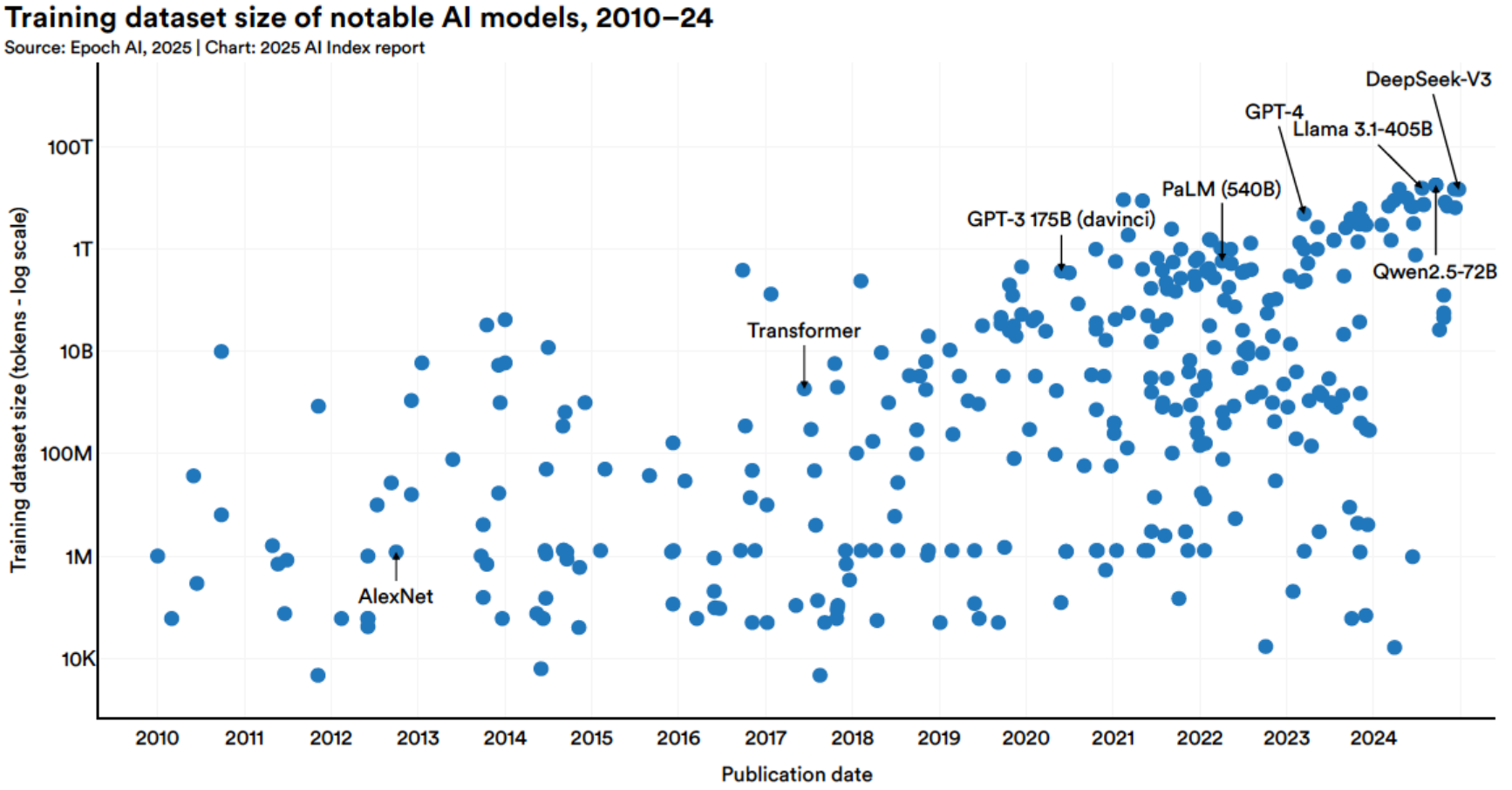


Figure 1.3.13

Credits: Artificial Intelligence Index Report 2025

What about raw data?

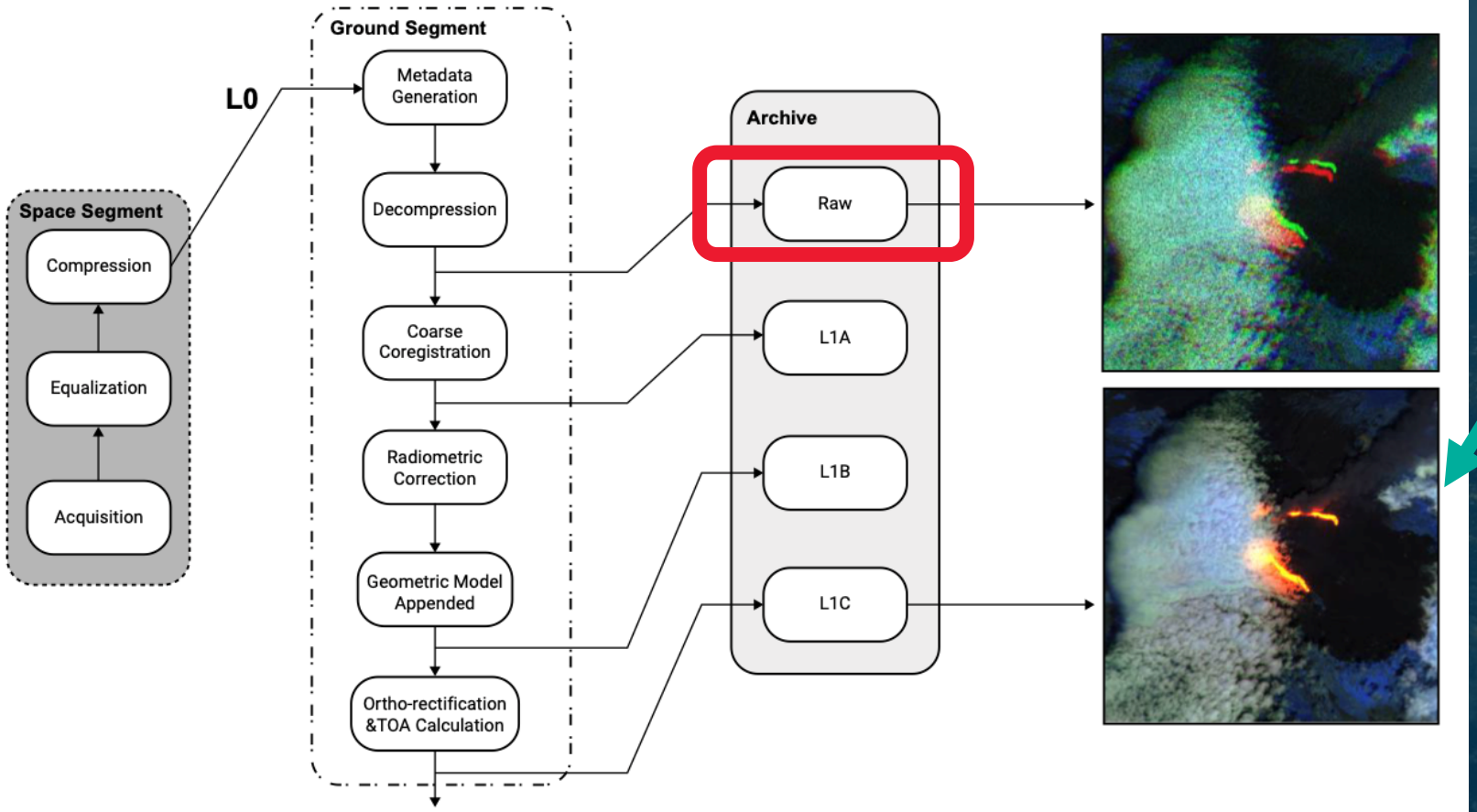


Fig. 1. Illustration of the processing chain from satellite data to L1C data. The images in the raw and L1C format show an eruption of the Etna volcano, Italy, for the two processing levels as RGB-like images.

Most of the GeoAI community focuses its attention at these data formats (L1C or higher)

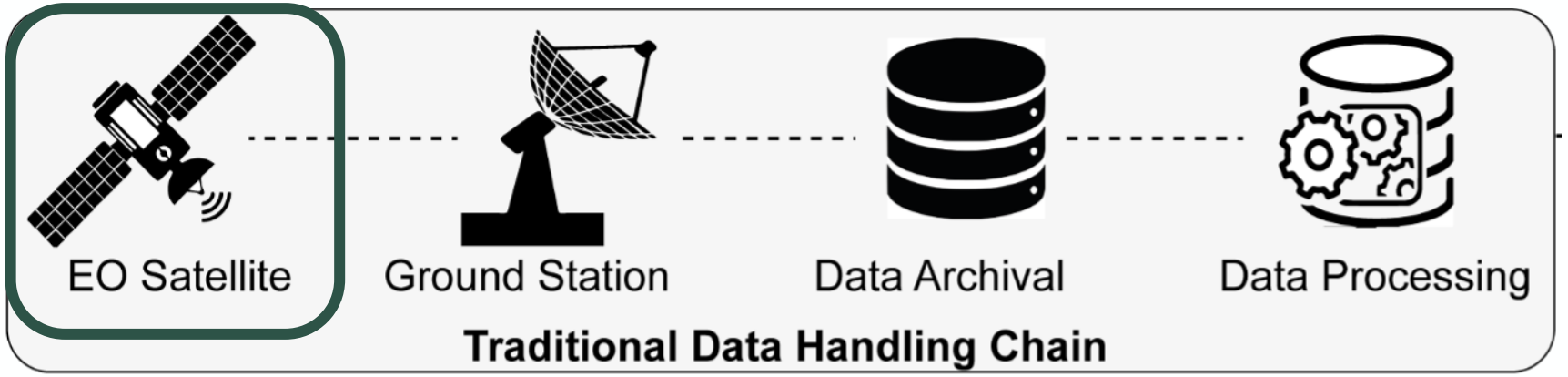
Question:

- Is there a better way to use all the raw data stored in the CDSE?
- Is there potential for new business cases?

Conventional ground processing

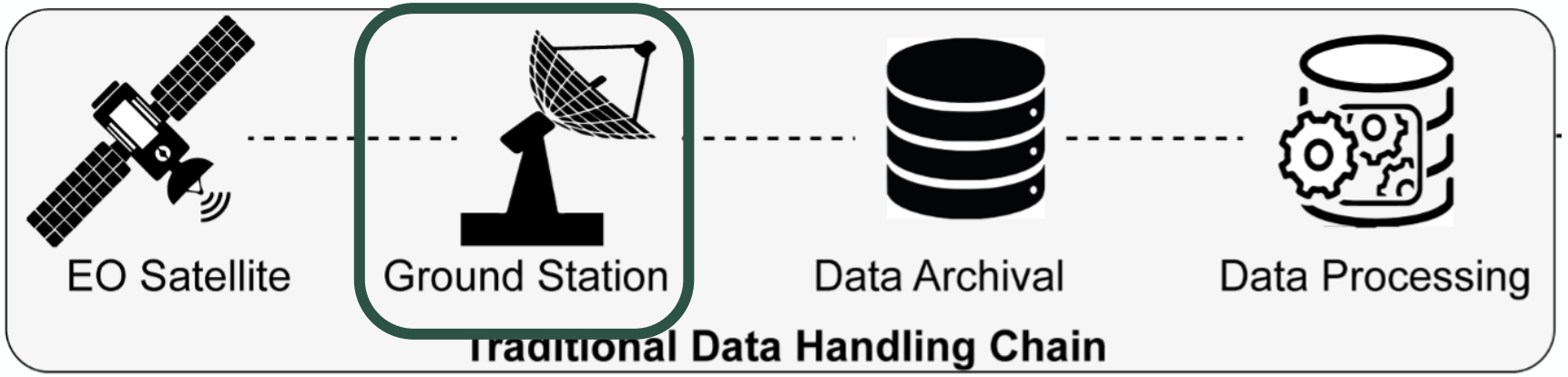
- **EO Satellite Acquisition**:
  - The satellite captures raw observation data (Level-0) using onboard sensors during its pass over the target area.
- **Downlink to Ground Station**:
  - The raw data is transmitted via radio frequency to a designated ground station during a contact window.
- **Ingestion and Archiving**:
  - The received data is ingested into the ground segment system and stored in a secure, structured
  - **Data Archive** for traceability and reprocessing.
- **Level-1A Processing**:
  - The raw data is radiometrically corrected and formatted into a standardized sensor grid (Level-1A), with no geometric correction applied yet.
- **Information Extraction**:
  - From Level-1A, higher-level products (e.g., L1B, L2) can be generated, enabling physical or thematic parameter retrieval (e.g., reflectance, vegetation index, land cover).

Conventional ground processing



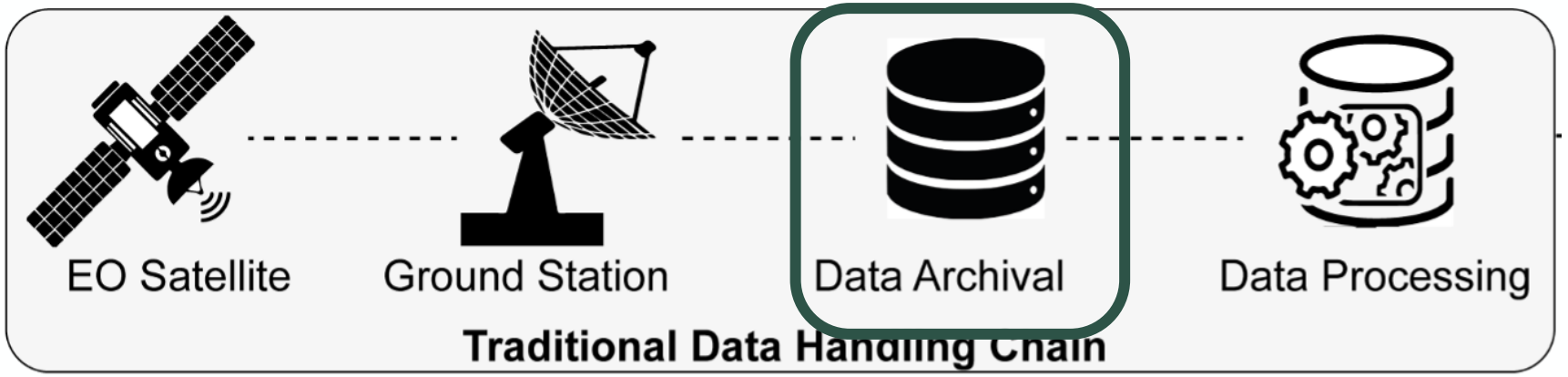
- **EO Satellite Acquisition:**
  - The satellite captures raw observation data (Level-0) using onboard sensors during its pass over the target area.
- **Downlink to Ground Station:**
  - The raw data is transmitted via radio frequency to a designated ground station during a contact window.
- **Ingestion and Archiving:**
  - The received data is ingested into the ground segment system and stored in a secure, structured
  - **Data Archive** for traceability and reprocessing.
- **Level-1A Processing:**
  - The raw data is radiometrically corrected and formatted into a standardized sensor grid (Level-1A), with no geometric correction applied yet.
- **Information Extraction:**
  - From Level-1A, higher-level products (e.g., L1B, L2) can be generated, enabling physical or thematic parameter retrieval (e.g., reflectance, vegetation index, land cover).

Conventional ground processing



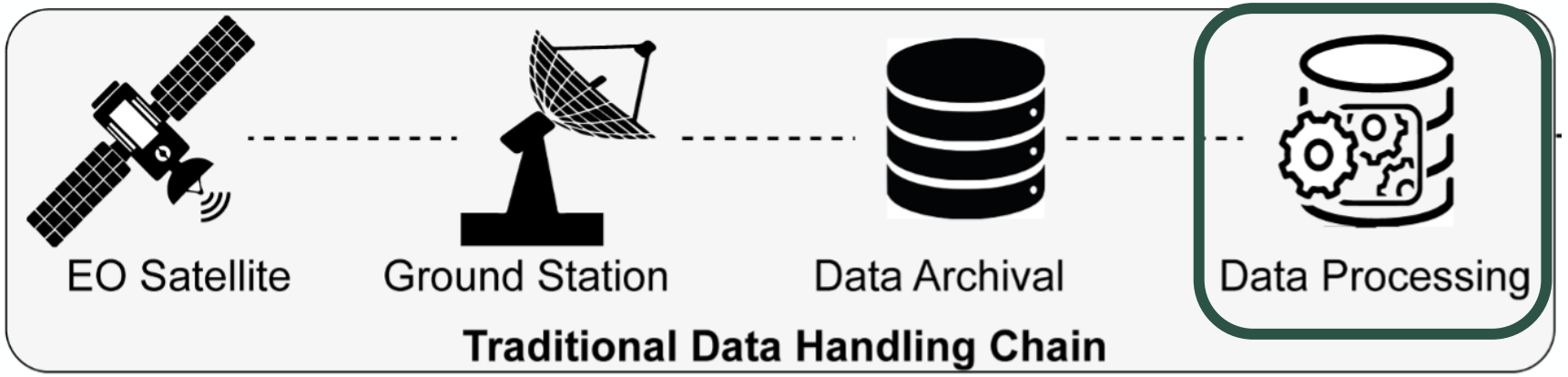
- EO Satellite Acquisition:
  - The satellite captures raw observation data (Level-0) using onboard sensors during its pass over the target area.
- **Downlink to Ground Station:**
  - The raw data is transmitted via radio frequency to a designated ground station during a contact window.
- Ingestion and Archiving:
  - The received data is ingested into the ground segment system and stored in a secure, structured
  - Data Archive for traceability and reprocessing.
- Level-1A Processing:
  - The raw data is radiometrically corrected and formatted into a standardized sensor grid (Level-1A), with no geometric correction applied yet.
- Information Extraction:
  - From Level-1A, higher-level products (e.g., L1B, L2) can be generated, enabling physical or thematic parameter retrieval (e.g., reflectance, vegetation index, land cover).

Conventional ground processing



- **EO Satellite Acquisition**:
  - The satellite captures raw observation data (Level-0) using onboard sensors during its pass over the target area.
- **Downlink to Ground Station**:
  - The raw data is transmitted via radio frequency to a designated ground station during a contact window.
- **Ingestion and Archiving**:
  - The received data is ingested into the ground segment system and stored in a secure, structured
  - **Data Archive** for traceability and reprocessing.
- **Level-1A Processing**:
  - The raw data is radiometrically corrected and formatted into a standardized sensor grid (Level-1A), with no geometric correction applied yet.
- **Information Extraction**:
  - From Level-1A, higher-level products (e.g., L1B, L2) can be generated, enabling physical or thematic parameter retrieval (e.g., reflectance, vegetation index, land cover).

Conventional ground processing



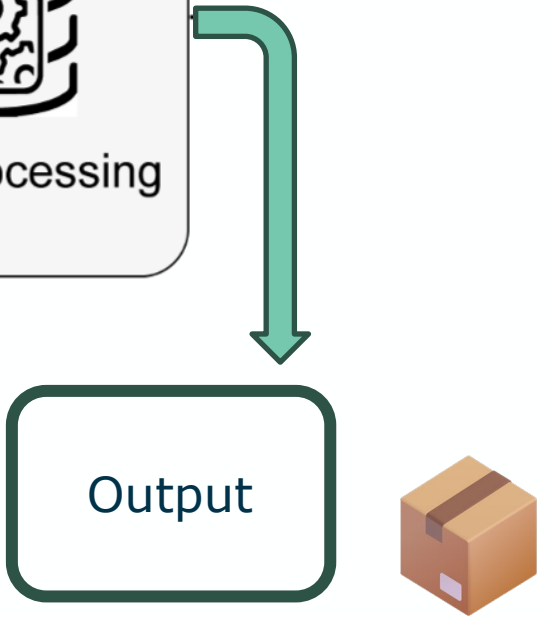
- EO Satellite Acquisition:
  - The satellite captures raw observation data (Level-0) using onboard sensors during its pass over the target area.
- Downlink to Ground Station:
  - The raw data is transmitted via radio frequency to a designated ground station during a contact window.
- Ingestion and Archiving:
  - The received data is ingested into the ground segment system and stored in a secure, structured
  - Data Archive for traceability and reprocessing.
- **Level-1A Processing**:
  - The raw data is radiometrically corrected and formatted into a standardized sensor grid (Level-1A), with no geometric correction applied yet.
- Information Extraction:
  - From Level-1A, higher-level products (e.g., L1B, L2) can be generated, enabling physical or thematic parameter retrieval (e.g., reflectance, vegetation index, land cover).

Conventional ground processing

- EO Satellite Acquisition:
  - The satellite captures raw observation data (Level-0) using onboard sensors during its pass over the target area.
- Downlink to Ground Station:
  - The raw data is transmitted via radio frequency to a designated ground station during a contact window.
- Ingestion and Archiving:
  - The received data is ingested into the ground segment system and stored in a secure, structured
  - Data Archive for traceability and reprocessing.
- Level-1A Processing:
  - The raw data is radiometrically corrected and formatted into a standardized sensor grid (Level-1A), with no geometric correction applied yet.
- **Information Extraction:**
  - From Level-1A, higher-level products (e.g., L1B, L2) can be generated, enabling physical or thematic parameter retrieval (e.g., reflectance, vegetation index, land cover).

Conventional ground processing

- **Only at this point the information can be delivered to the end-user.**



Is this approach the best for every application? . . .

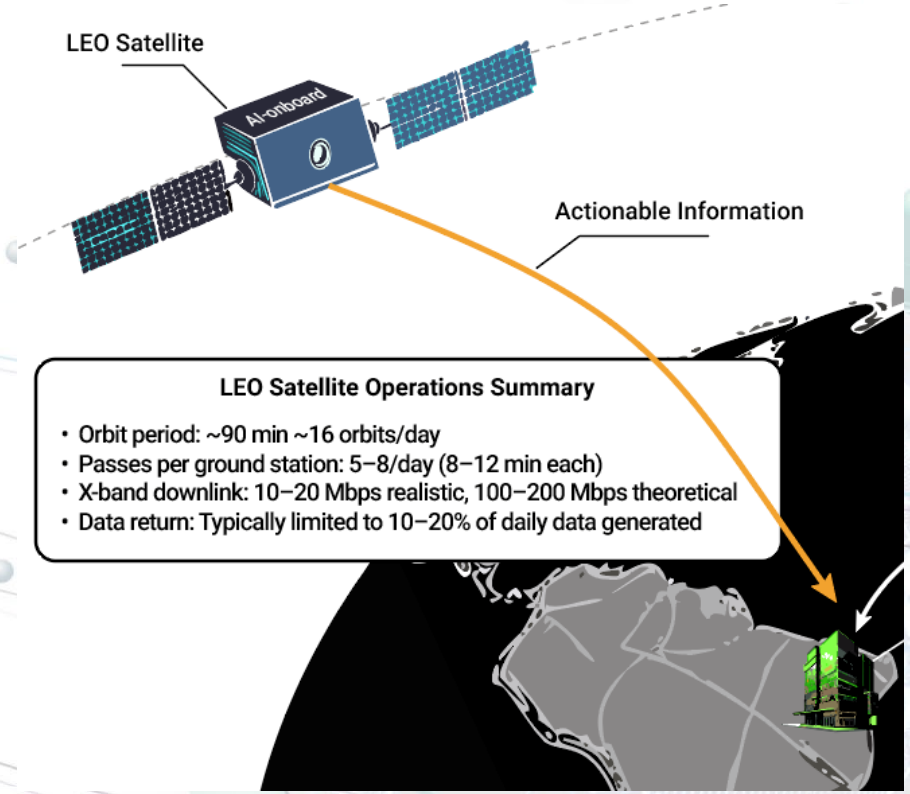
Conventional ground processing - Challenges

Data Latency:
- Time lag between acquisition, downlink, processing, and delivery to end-users.
- Critical bottleneck for time-sensitive applications like emergency response.

Data Volume:
- Modern EO missions (e.g., Sentinel, Landsat, PlanetScope) generate petabytes of multi-sensor, multi-temporal data.
- Requires high-throughput processing pipelines and scalable storage solutions.

Ground Segment Bottlenecks:
- Traditional centralized ground-based processing can become overwhelmed, especially during high-demand periods (e.g., disaster events).

👉 For some applications, we need effective ways to deliver QUICKLY the information gathered by EO sats



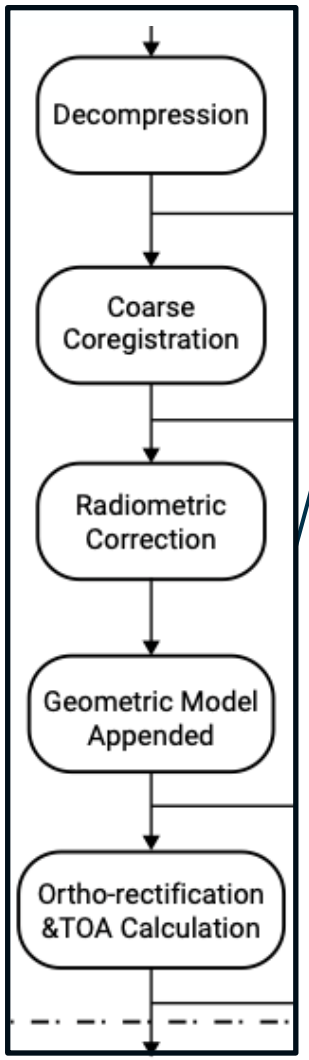
Onboard AI Missions

NEW SPACE

Onboard Autonomy

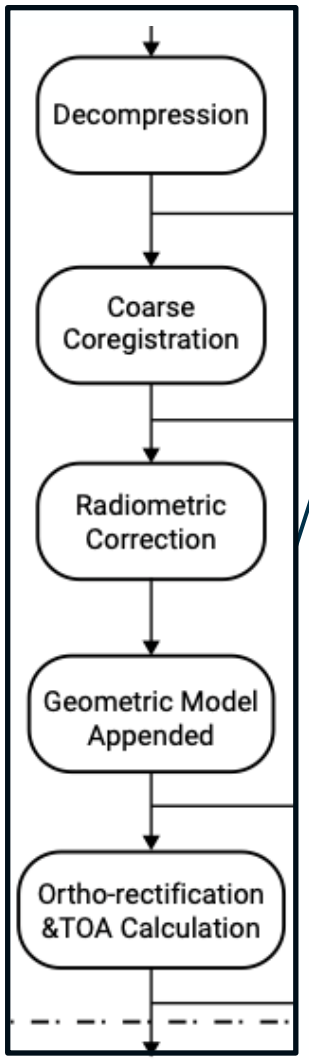
Vertical Segment specialized on specific applications. Need for low power computing on orbit

Data Handling



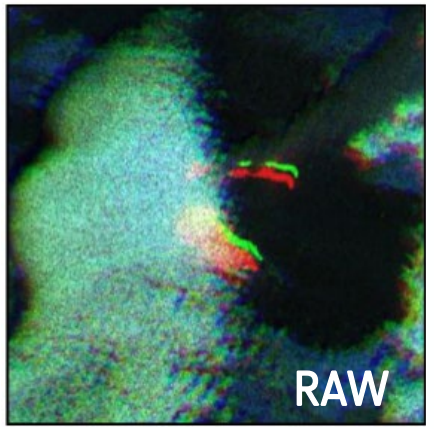
On-orbit replication of this pipeline is computationally heavy, requiring specialized hardware.

Onboard Artificial Intelligence



On-orbit replication of this pipeline is computationally heavy, requiring specialized hardware.

Idea:
Explore use of lower-level products (e.g., Level-0/1A) to reduce onboard processing cost and enable direct, near-real-time feature extraction.



AI Model

OUTPUT

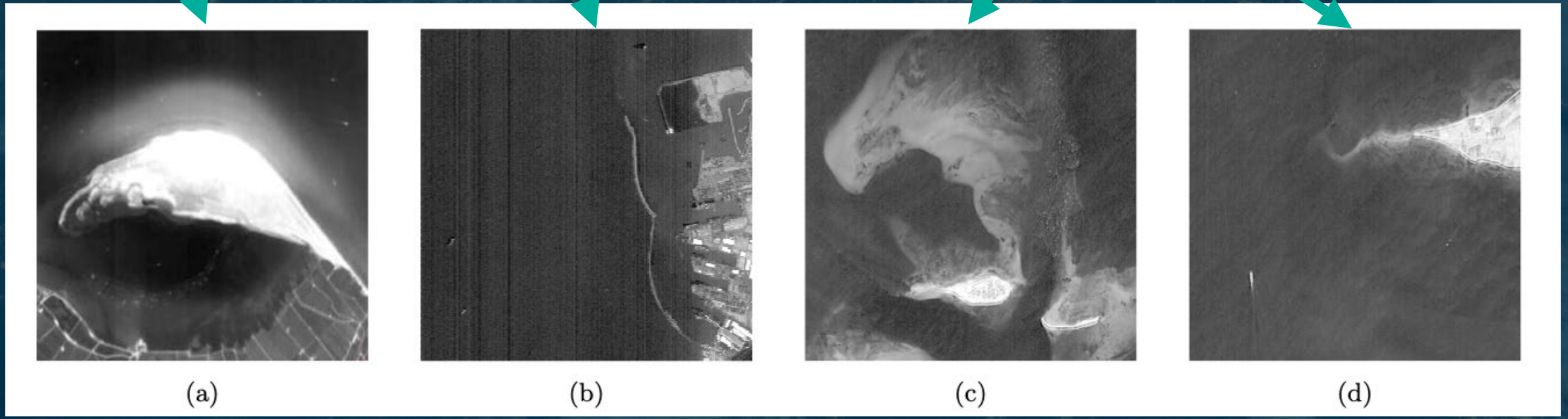
Actionable Information
Real Value

Unique information in Raw Data

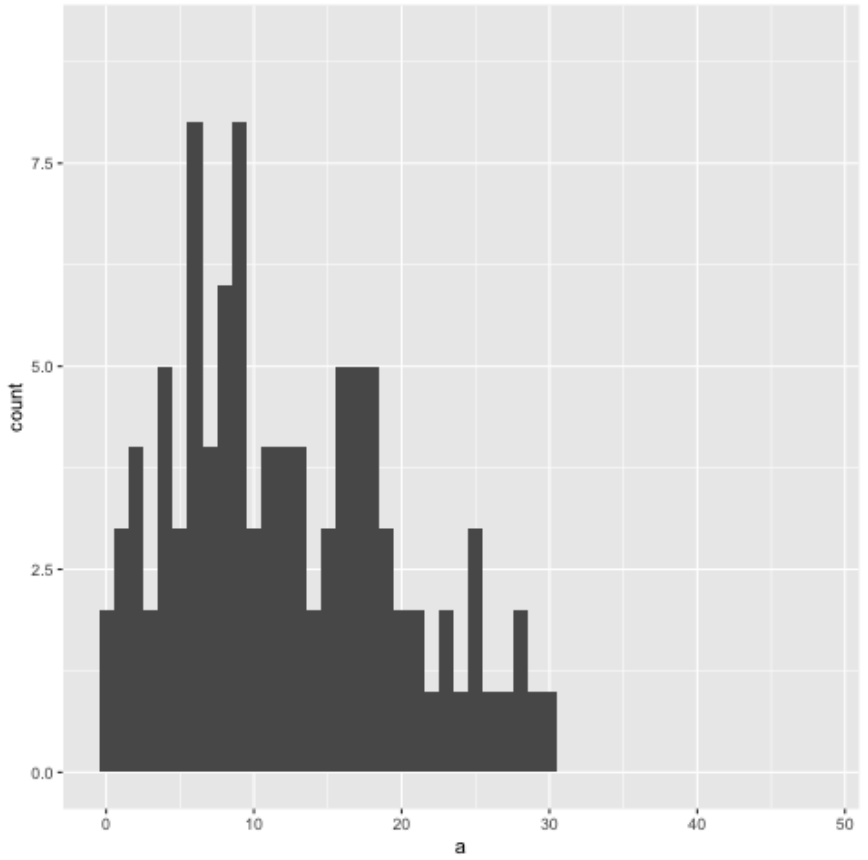
a) Stray lights artifacts

b) Striping Noise

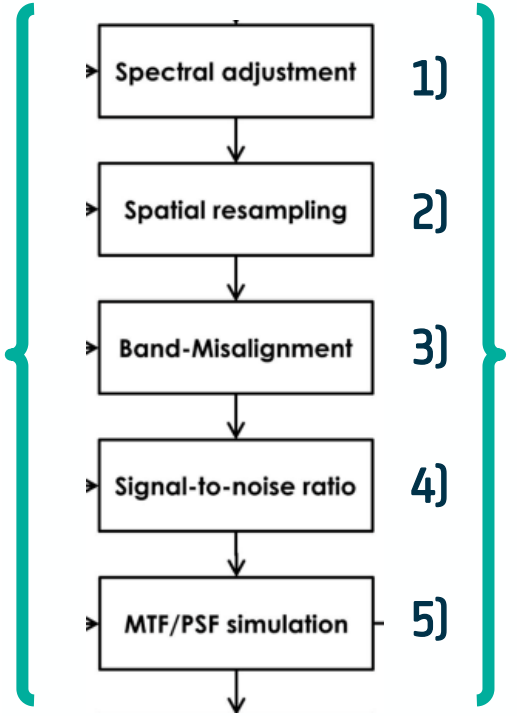
c) Radiometric Noise



Unique information in Raw Data



L0 and L1 are not the same..

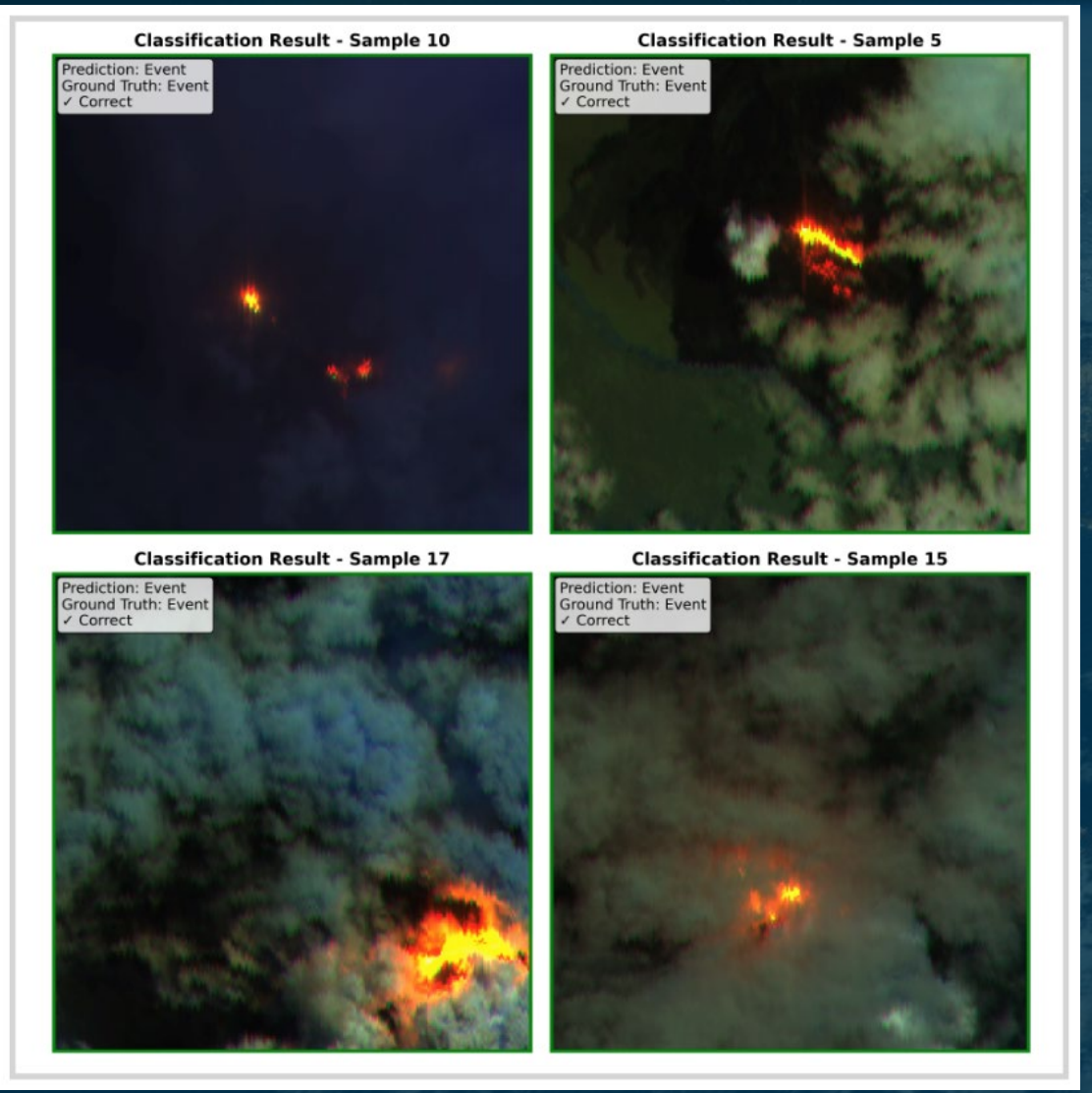


We can roughly summarize 5 main points of difference between L0 and L1.

↓

Simulating L0 from L1 can yield incorrect results if L1 lacks the expressive power or closure properties needed to capture L0's structure.

Potential Applications



THRawS

SCAN ME

G. Meoni, R. D. Prete, F. Serva, A. De Beusscher, O. Colin and N. Longépé, "Unlocking the Use of Raw Multispectral Earth Observation Imagery for Onboard Artificial Intelligence," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 17, pp. 12521-12537, 2024, doi: 10.1109/JSTARS.2024.3418891.
keywords: {Artificial intelligence;Radiometry;Satellite broadcasting;Metadata;Image coding;Earth;Pipelines;Onboard artificial intelligence (AI);raw dataset;Sentinel-2;volcanic eruption;wildfire},

- THRawS is a global dataset of raw Sentinel-2 data featuring warm temperature hotspots like wildfires and volcanic eruptions.
- It supports developing energy-efficient preprocessing algorithms and AI models for onboard satellite use.
- A custom method links raw data to Level-1C products using coarse coregistration, with comparisons between the proposed technique and SuperGlue for spatial accuracy.

Potential Applications

VDS2Raw

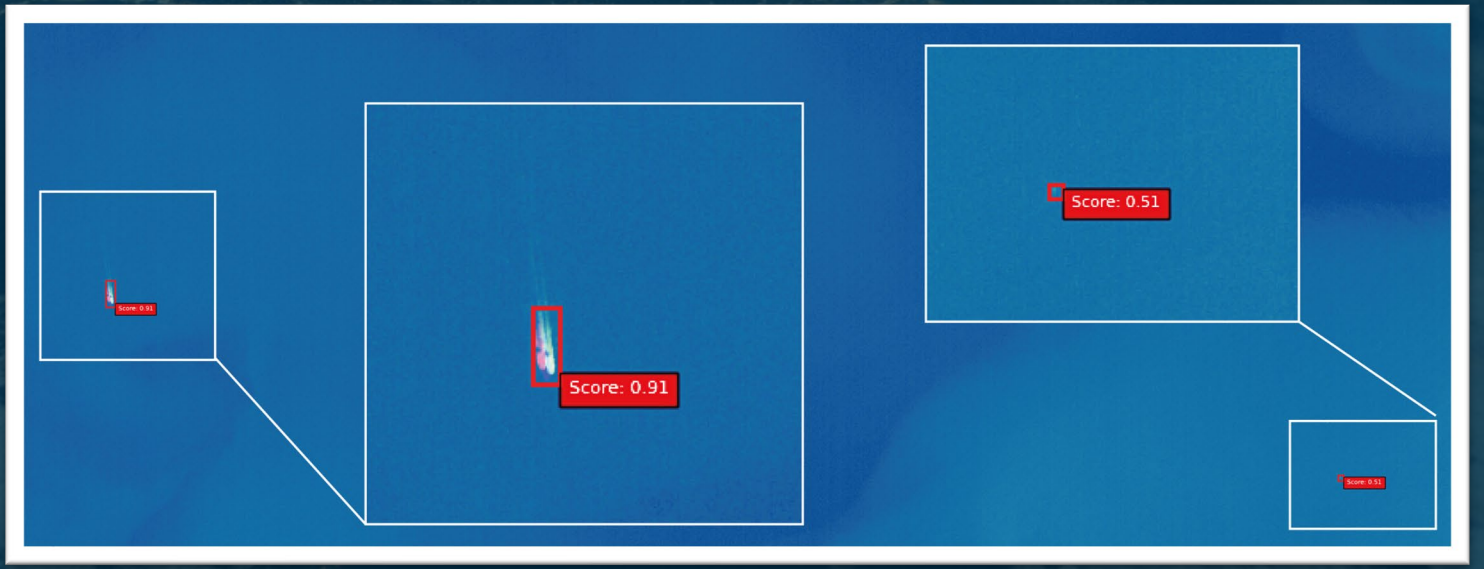
R. Del Prete et al., "Enhancing Maritime Situational Awareness Through End-to-End Onboard Raw Data Analysis," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 18, pp. 16997-17018, 2025, doi: 10.1109/JSTARS.2025.3584999.
keywords: {Artificial intelligence;Satellite broadcasting;Radiometry;Real-time systems;Space missions;Sensors;Meters;Hardware;European Space Agency;Earth;Embedded systems;raw multispectral data;Sentinel-2 (S-2);Vegetation and Environment Monitoring New Micro Satellite (VEN μS);vessel detection},

SCAN ME

390 Level-0 granules were processed with ESA's PyRaws tool and manually annotated across bands B02, B03, B04, and B8.

Composition:

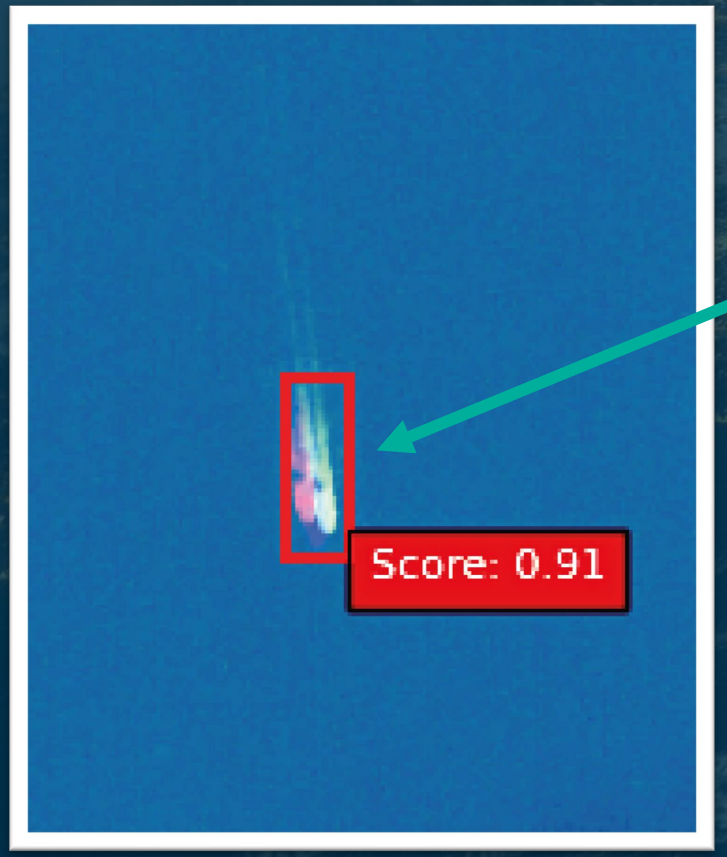
166 curated granules (105 train / 27 val / 34 test) include 695 ships, with average image size ~2589×1669 pixels and mean ship size ~13.6×15.7 pixels.



Potential Applications

VDS2Raw

R. Del Prete et al., "Enhancing Maritime Situational Awareness Through End-to-End Onboard Raw Data Analysis," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 18, pp. 16997-17018, 2025, doi: 10.1109/JSTARS.2025.3584999. keywords: {Artificial intelligence;Satellite broadcasting;Radiometry;Real-time systems;Space missions;Sensors;Meters;Hardware;European Space Agency;Earth;Embedded systems;raw multispectral data;Sentinel-2 (S-2);Vegetation and Environment Monitoring New Micro Satellite (VEN μ S);vessel detection},



AI can cope with coarsely co-registered bands and still produce valuable insights.

→ Bypass computationally demanding processing onboard.

SCAN ME

Concluding remarks

Untapped potential of RAW data

- Large volumes of raw EO (Earth Observation) data remain underexploited.
- Direct access to unprocessed data allows for novel insights, higher temporal fidelity, and improved model interpretability.
- RAW-level analytics enable adaptation to diverse user needs without dependency on predefined processing chains.

Concluding remarks

Untapped potential of RAW data

- Large volumes of raw EO (Earth Observation) data remain underexploited.
- Direct access to unprocessed data allows for novel insights, higher temporal fidelity, and improved model interpretability.
- RAW-level analytics enable adaptation to diverse user needs without dependency on predefined processing chains.

AI performance depends on data quality and accessibility

- AI models require consistent, high-quality, and representative datasets to achieve operational robustness.
- Broader access to calibrated, well-documented raw data is essential to improve domain adaptation and reduce model bias.
- The community is increasingly advocating for open and standardized access to such datasets to support innovation and reproducibility.

Concluding remarks

Untapped potential of RAW data

- Large volumes of raw EO (Earth Observation) data remain underexploited.
- Direct access to unprocessed data allows for novel insights, higher temporal fidelity, and improved model interpretability.
- RAW-level analytics enable adaptation to diverse user needs without dependency on predefined processing chains.

AI performance depends on data quality and accessibility

- AI models require consistent, high-quality, and representative datasets to achieve operational robustness.
- Broader access to calibrated, well-documented raw data is essential to improve domain adaptation and reduce model bias.
- The community is increasingly advocating for open and standardized access to such datasets to support innovation and reproducibility.

Business opportunities through raw data exploitation

- Vertical markets (e.g., agriculture, energy, maritime, defense) can extract greater value by integrating analytics directly at the data source.
- Edge-AI deployments can process raw data locally, reducing bandwidth usage, latency, and dependency on centralized infrastructure.
- On-site inference from raw sensor streams opens pathways for near-real-time decision-making and customized commercial services.

The Copernicus Data Space Ecosystem stands at a pivotal point to unlock the latent value of **RAW** data.

By enabling standardized, scalable access and promoting AI-ready datasets at the source, CDSE can catalyze a new generation of edge intelligence applications.

This shift will empower both research and industry to build faster, leaner, and more adaptive Earth observation solutions—positioning Europe at the forefront of data-driven innovation.

Thanks for the attention