Earth Observation & Weather Data Federation with AI Embeddings

D5.3: Standardisation, Exploitation and Sustainability Strategy and Plan

Revision: v 2.0

DX.X: Deliverable Title

Sub-title here if needed/appropriate

Work package	WP5
Task	Task 5.3
Due date	10/October/2025
Submission date	10/October/2025
Deliverable lead	IBM Research
Version	2.0
Authors	Thomas Brunschwiler, Martel, UOXF, SATCEN, HISDESAT
Reviewers	Michele Lazzarini (SATCEN)
Abstract	To maximize the impact of Embed2Scale, the exploitation plan is structured around three core objectives: • First, it establishes a comprehensive inventory of all exploitable assets generated by the project, including software tools, datasets, models, methodologies, and know-how. • Second, it identifies key stakeholder groups such as industry partners, public institutions, and the research community, that stand to benefit from these assets, ensuring targeted and relevant dissemination. • Third, it outlines the go-to-market strategies both at the individual partner level and for the consortium as a whole, taking into account varied commercialization pathways, licensing models, and collaboration opportunities to drive sustainable uptake and long-term value creation.

Document Revision History

Version	Date	Description of change	List of contributor(s)
V0.1	01/07/2025	Partner interviews	Thomas Brunschwiler (IBM) Conrad Albrecht (DLR) Roya Shokoohi (Martel Innovate BV) Jonas Hurst (Universität Münster) Matej Batič (Sinergise Solutions) Tim Reichelt (University of Oxford) Rocco Sedona (FZJ)
V1.0	31/07/2025	Deliverable writing	Thomas Brunschwiler (IBM)
V2.0	10/10/2025	Response to reviewers comments	Thomas Brunschwiler (IBM)

Grant Agreement No: 101131841 | **Topic:** HORIZON-EUSPA-2022-SPACE-02-55
Call: HORIZON-EUSPA-2022-SPACE | **Type of action:** HORIZON-RIA

DISCLAIMER

Embed2Scale (Earth Observation & Weather Data Federation With AI Embeddings) project is funded by the EU's Horizon Europe program under Grant Agreement number 101131841. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them. This work has also received funding from the Swiss State Secretariat for Education, Research and Innovation (SERI) and the UK Research and Innovation (UKRI).

COPYRIGHT NOTICE

Project funded by the European Commission in the Horizon Europe Programme		
Nature of the deliverable:	R	
Dissemination Level		
PU	Public, fully open, e.g. web (Deliverables flagged as public will be automatically published in CORDIS project's page)	

TABLE OF CONTENTS

SUMMARY

To maximize the impact of Embed2Scale, the exploitation plan is structured around three core objectives:

- First, it establishes a comprehensive inventory of all exploitable assets generated by the project, including software tools, datasets, models, methodologies, and know-how.
- Second, it identifies key stakeholder groups such as industry partners, public institutions, and the research community, that stand to benefit from these assets, ensuring targeted and relevant dissemination.
- Third, it outlines the go-to-market strategies both at the individual partner level and for the consortium as a whole, taking into account varied commercialization pathways, licensing models, and collaboration opportunities to drive sustainable uptake and long-term value creation.

1 STANDARDIZATION, EXPLOITATION AND SUSTAINABILITY PLAN

OBJECTIVE OF THE STANDARDIZATION, EXPLOITATION AND SUSTAINABILITY PLAN

The objective of this Standardization, Exploitation, and Sustainability Plan is to ensure that the innovations and outcomes of the **Embed2Scale** project are not only effectively leveraged during the project's lifetime but also translated into long-term impact beyond its completion. This deliverable outlines the strategic approach taken to identify, refine, and implement pathways for exploiting the project's results, securing their sustainability, and aligning them with relevant standards and market expectations. Led by **Task 5.3,** and closely integrated with the technical and validation activities of **Work Packages 1–4**, the plan represents a critical step in ensuring that the project's results reach their full potential in terms of commercial, societal, and environmental value.

The plan encompasses a range of actions (Figure 0), including the identification of **exploitable results, market assessments, and value propositions** tailored to the key outcomes of Embed2Scale. By incorporating early exploitation planning and stakeholder engagement, including internal workshops, external consultations, and piloting feedback, the project ensures a robust understanding of market dynamics and user needs.

Additionally, this deliverable presents the foundation for a joint **Exploitation Approach** among partners, supporting both collaborative and individual exploitation pathways while addressing intellectual property rights and strategic roles. Crucially, the sustainability component ensures that key results remain viable and supported after project completion, through targeted activities, partnerships, and market positioning.

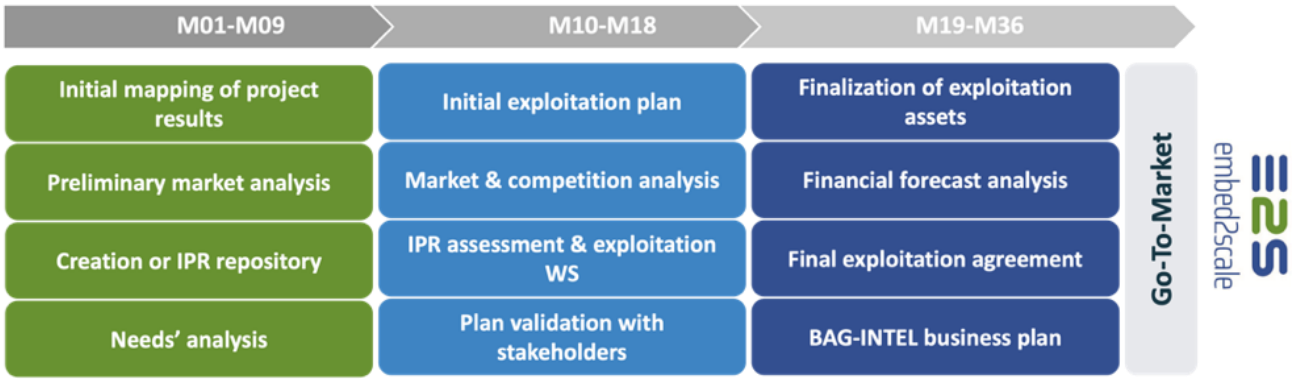


Figure 0: Depicting the three exploitation phases in the project: i) Awareness creation and communication foundation, ii) Community building, initial results dissemination, and iii) Global outreach, exploitation and sustainable impact, continued by the go-to-market integration beyond the project

This plan plays a central role in bridging project outcomes with real-world application and longevity. It defines clear steps for value creation, impact maximization, and enduring relevance, ensuring that **Embed2Scale** innovations contribute meaningfully to both the targeted sectors and the broader European innovation ecosystem.

2 PROJECT OBJECTIVES

PROJECT VISION

The Embed2Scale project envisions a transformative approach to **Earth Observation (EO)** and **climate data** processing by enabling highly efficient, scalable, and interoperable AI technologies. At its core, the project aims to compress and structure vast amounts of EO data into actionable, lightweight embeddings that can be easily transferred, interpreted, and reused across platforms and applications. This vision is realized through four interlinked technological innovations:

- **Embed2Transfer:** Reducing data gravity by enabling decentralized applications through efficient embedding-based data abstraction, allowing data to stay where it is while the intelligence moves.
- **Embed2Infer:** Enhancing the portability of inference by minimizing computational demand, enabling lightweight model execution on edge and low-power environments.
- **Embed2Train:** Accelerating training processes by using few-shot model architectures, thereby reducing the need for large volumes of labeled data.
- **Embed2Find:** Enabling near-real-time similarity search across petabyte-scale data archives, fostering scalable knowledge discovery and rapid insight generation.

PROJECT OBJECTIVES

To fulfil this vision, Embed2Scale is guided by four high-level objectives, each backed by specific technological targets and deliverables.

Objective 1: Explore groundbreaking AI-compressors to enable data federation via embeddings

1.1 Development of AI-compressor architectures producing highly compressed embeddings using generative self-supervised learning.
1.2 Design of AI-encoders that generate embeddings requiring minimal computation for downstream inference tasks.
1.3 Creation of embeddings optimized for training downstream models with few-shot learning capabilities.
1.4 Integration of high-speed vector search algorithms to enable real-time retrieval of scene similarities from stored embeddings.

Objective 2: Proliferate ML-Ops reference implementation for embedding generation on HPC clusters

2.1 Deployment of a standardized ML-Ops reference implementation on partner high-performance computing (HPC) clusters.
2.2 Implementation of a System-Level Life-Cycle Model to evaluate the energy efficiency and performance impact of downstream applications with and without Embed2Scale technology.

Objective 3: Demonstrate data federation in real-world use-cases using EO and climate data

3.1 Use-case on Maritime Awareness, showcasing detection and tracking capabilities in distributed settings.
3.2 Estimation of Aboveground Biomass for biodiversity monitoring and carbon accounting.
3.3 Prediction of Climate and Air Pollution trends using federated inference pipelines.
3.4 Detection of Crop Stress and Early Yield, leveraging compressed embeddings to inform precision agriculture.

Objective 4: Stimulate community acceptance and contribution through open-source and standards

4.1 Release of the SSL4EO-S12 dataset under open-source terms, along with an initial benchmarking campaign.
4.2 Integration of Embed2Scale functionality within the OpenEO ecosystem to support broad interoperability.
4.3 Exposure of the Embed2Scale API to the Open Geospatial Consortium (OGC) to align with international data standards.

The consortiums decision on actions to make progress towards these objectives will be traced in the meeting minutes of the recuring Embed2Scale consortium meetings.

3 PROJECT ASSETS

ASSET INVENTORY

Within Embed2Scale, several assets are created according to the mentioned project objectives. They can impact the market, the community and society as individual assets or in combination thereof. An overview of the inventory is provided in Table 1. Details of the assets are discussed in the subsequent sub-chapters. In the final exploitation deliverable (D5.4), we will also provide quantitative KPIs (e.g. stars, down-loads, user contributions) for the various assets listed in the asset inventory.

Table 1: Overview of the Embed2Scale assets

Asset Name	Asset Type	Work Package & Objective	License
CDSE Embeddings[1]	Embed2Scale embeddings hosted on Copernicus Data Space Ecosystem	WP-1 Objective-4	CC-BY-SA-4.0
NeuCo-Bench[2]	Benchmarking framework to evaluate lossy compression methods in earth observation	WP-2 Objective-1&4	Apache License 2.0
Climate-Bench-Press	Benchmarking framework to evaluate lossy compression for weather and climate data	WP-2 Objective-1&4	GitHub Mozilla Public License 2.0
Compressor-Models	Neural compressor models dedicated to various earth observation modalities	WP-2 Objective-1	Apache License 2.0
TerraTorch-Embeddings[3]	Python package to enable embedding generation by TerraTorch	WP-3 Objective-2	Apache License 2.0
openEO-embeddings[4]	API extension for openEO and STAC catalogue to manage embeddings, incl. federation	WP-3 Objective-3&4	Apache License 2.0

[1] https://dataspace.copernicus.eu/

[2] https://github.com/embed2scale/NeuCo-Bench

[3] https://github.com/IBM/terratorch

[4] https://github.com/open-eo

Embedding-Energy-Monitor	Library to quantify energy and environmental impact of embedding generation	WP-3 Objective-2	Apache License 2.0
Vessel-Detector	AI-model to detect vessels from satellite observations	WP-4 Objective-3	Proprietary
Embedding Workshop Material	Embed2Scale Workshop Material for internal and external use	WP-5 Objective-4	n.a.
Earth2Vec Community	E2S initiated, global community pooling major players exploring EO embeddings	WP-5 Objective-4	n.a.

CDSE EMBEDDINGS

The Copernicus Data Space Ecosystem (CDSE) is a comprehensive digital infrastructure that provides open and free access to the vast amount of Earth observation data collected by the Copernicus programme, particularly from Sentinel satellites. Launched by the European Union, it serves as a central hub for accessing, processing, and distributing geospatial data, enabling users to explore real-time and historical environmental information. The ecosystem includes cloud-based tools, APIs, and services designed for researchers, policymakers, developers, and commercial users alike.

Its importance lies in promoting sustainability, innovation, and informed decision-making across a range of sectors such as climate monitoring, disaster response, agriculture, urban planning, and environmental protection. By lowering the barrier to entry for satellite data usage, the Copernicus Data Space Ecosystem fosters collaboration, supports the digital transition in Europe, and strengthens global capabilities to address pressing environmental and societal challenges.

CloudFerro operates the CDSE and as part of their innovation efforts has published global and dense embeddings of earth based on the Major TOM data set[5], enabling advanced search, analysis, and machine learning applications on Copernicus Sentinel data.

IBM Research is partnering with CloudFerro to generate and publish compressed embeddings, by the Embed2Scale Compressor Models and the TerraTorch Embedding library. The compressed embeddings will be available on CDSE, with demos demonstrating 100x inference and 100x model training, compared to state-of-the-art models.

NEUCO-BENCH

NeuCo-Bench is a benchmarking framework for assessing how well compressed embeddings retain information relevant to downstream tasks in data-intensive domains like Earth Observation (EO) (Figure 1).

[5] Czerkawski, Mikolaj, Marcin Kluczek, and J. Ä. Bojanowski. "Global and dense embeddings of earth: Major tom floating in the latent space." *arXiv preprint arXiv:2412.05600* (2024).

While traditional compression prioritizes pixel-level fidelity and Foundation Models (FMs) often ignore embedding size, NeuCo-Bench bridges this gap by enforcing strict embedding size limits and directly evaluating their utility in real-world EO applications. The framework currently includes a suite of EO tasks, namely compiled from the Embed2Scale applications, such as land-cover proportion estimation, cloud detection, and biomass estimation and encourages the community to contribute additional tasks and datasets from EO and beyond.

Framework Highlights

- Model-Agnostic: Compatible with any fixed-size embedding (e.g., 1024-dimensional vectors), allowing fair comparisons across both compression and representation learning methods.
- Task-Driven Evaluation: Uses linear probes on diverse EO tasks to assess embedding quality.
- Robust Metrics: Employs signal-to-noise ratios and dynamic rank aggregation for comprehensive method comparison.

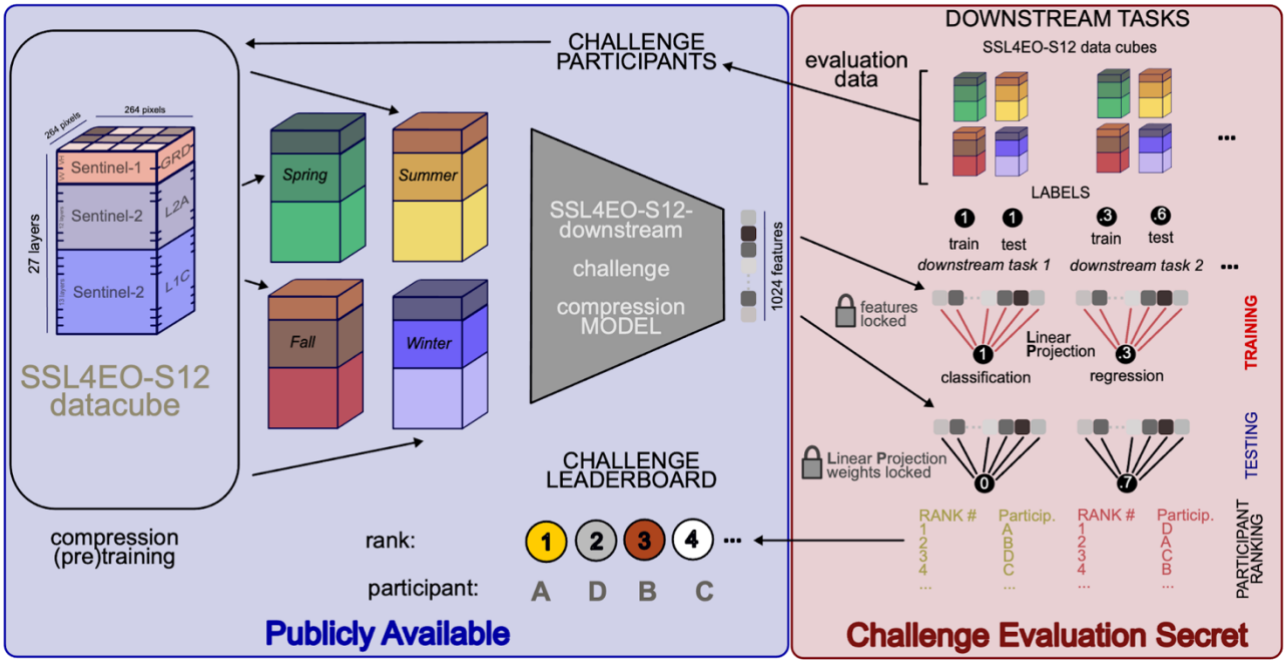


Figure 1: NeuCo-Bench overview image.

CLIMATE-BENCH- PRESS[6]

ClimateBenchPress is a benchmark suite designed to evaluate and advance lossy compression methods for weather and climate data. As high-resolution simulations increasingly generate data at rates exceeding 20 GB per second, efficient compression becomes essential for storage, analysis, and sharing. ClimateBenchPress addresses this need by providing standardized datasets and

[6] Reichelt, T., Tyree, J., Kloewer, M., Dueben, P., Lawrence, B., Hammerling, D., Baker, A., Faghih-Naini, S., and Stier, P.: ClimateBenchPress: A Benchmark for Compression of Climate Data, EGU General Assembly 2025, Vienna, Austria, 27 Apr–2 May 2025, EGU25-15912, https://doi.org/10.5194/egusphere-egu25-15912, 2025.

evaluation protocols to test the performance and scientific fidelity of lossy compressors. The benchmark spans atmospheric, oceanic, and terrestrial data with diverse spatial (1–150 km) and temporal (hourly to yearly) resolutions, and includes tasks ranging from stable single-variable fields to large, multi-variable archives with evolving statistics. Compressors are scored based on their ability to minimize file size without compromising scientific integrity, with tests designed to reflect real-world use cases. Baseline results using SZ3, ZFP, and Real Information are included for comparison, establishing reference points and supporting the broader goal of safe, standardized climate data compression. The library will be accessible open-source on GitHub Mozilla Public License 2.0.

COMPRESSOR- MODELS

A competitor analysis (e.g. MajorTOM, Prithvi and Google embeddings) of Earth Observation embeddings was performed and depicted in a spatio-temporal overivew (Figure 2). In Embed2Scale, we aim at patch and image level embedding in an initial stage for single time steps, while extending it to bi- and multi-temporal embeddings over time.

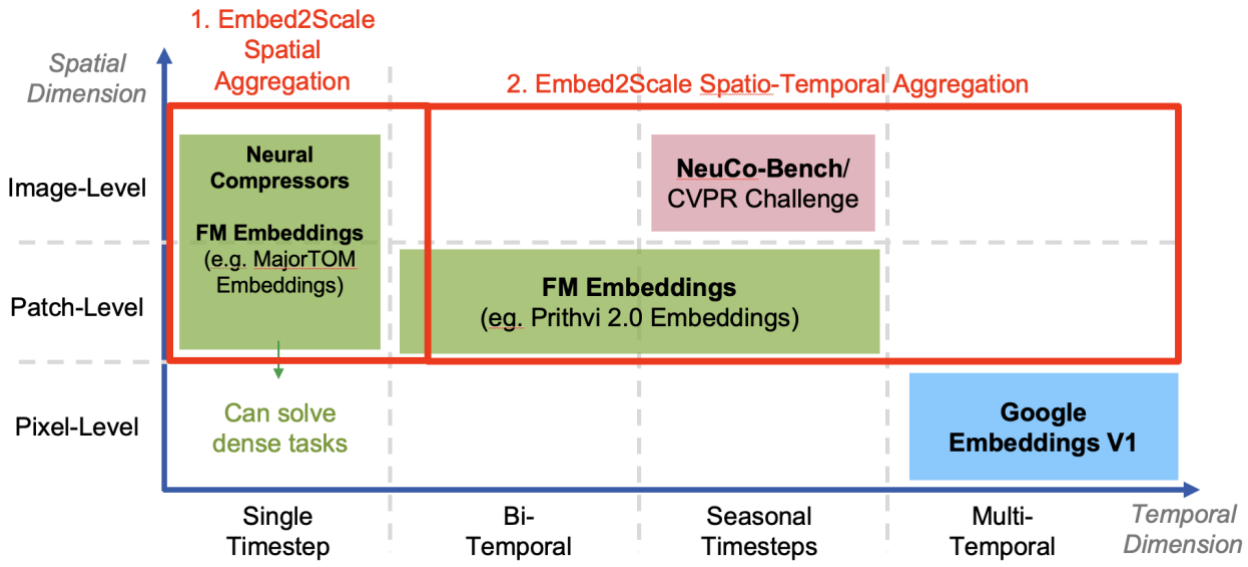


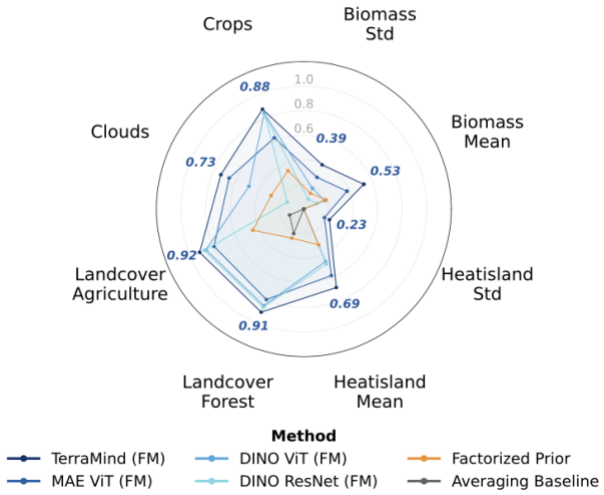
Figure 2: Overview of embedding landscape and focus of Embed2Scale Compressor Models.

The compression methods developed in Embed2Scale will be integrated into TerraTorch to generate the compressed embeddings based on the following methods:

i) Baseline compression means are applied to foundation model embeddings. A simple averaging baseline performs spatial downsampling, channel-wise averaging, and flattening.

ii) More advanced methods use Factorized Prior autoencoders pretrained on SSL4EO data, extracting and downsampling latent bottleneck features before entropy coding.

iii) TerraMind[7] tokens: TerraMind is a multimodal Earth Observation Foundation Model developed by IBM Research in the ESA FAST-EO project. In Embed2Scale, we use the tokenizers to result in efficient compression (Figure 3).



Method	Pretraining	Loss	Backbone	Input	Downsampling	Original latent dim.
TerraMind (FM)	TerraMesh	Cross-Entropy	Multi-modal ViT-B/16	All (S1,S2)	Patch averaging	196 × 768
MAE ViT (FM)	SSL4EO	MAE	ViT-B/16	S2L1C	Patch averaging	196 × 768
DINO ViT (FM)	SSL4EO	DINO	ViT-B/16	S2L1C	Patch averaging	196 × 768
DINO ResNet (FM)	SSL4EO	DINO	ResNet-50	S2L1C	Global avg., pairwise channel mean	2048 × 7 × 7
Factorized Prior	SSL4EO	Rate-Distortion	CNN	S2L1C	Adaptive pool (4 × 4), pairwise channel mean	128×14×14
Averaging Baseline	–	–	–	All (S1,S2)	bilinear avg., flatten	4 × 8 × 8

Figure 3: TerraMind embeddings outperform all other models on the NeuCo-Bench benchmark.

iv) Temporal Neural Compression for Earth Observation introduces a domain-specific neural codec tailored to compress multispectral satellite time series efficiently. Unlike traditional image or video codecs, the model leverages a transformer-based architecture (an extension of the Video Compression Transformer) to capture EO-specific temporal patterns like seasonal changes and cloud dynamics. A key innovation is the adaptive token mechanism, which allows dynamic control over compression quality, enabling a single model to serve diverse use cases without retraining.

TERRATORCH- EMBEDDINGS

TerraTorch is an open-source toolkit from IBM Research designed to streamline the use of Earth Observation Foundation Models (EO-FMs) for extracting high-level representations from satellite imagery. Built on PyTorch, it provides a unified interface to a curated set of pretrained EO-FMs, including both unimodal (e.g., optical-only) and multimodal (e.g., radar-optical) architectures such as ResNet, ViT, and TerraMind. TerraTorch supports various data formats commonly used in Earth observation (e.g., GeoTIFF, JPEG2000, Zarr), automatically handling preprocessing steps like normalization, resampling, and band selection. By abstracting away model-specific details and data handling, it lowers the barrier for researchers and practitioners to apply cutting-edge models to new datasets or operational pipelines.

Within Embed2Scale, we will extend TerraTorch to generate compressed embeddings. Users simply load their imagery, select a pretrained model, and extract fixed-size feature vectors from a specified layer, just before the classification head with the additional Compressor Model codex. These embeddings can then be published and used in a wide range of downstream applications enabling Embed2Find, Embed2Transfer, and Embed2Infer. TerraTorch also supports batched processing and GPU acceleration, enabling efficient large-scale embedding generation for EO archives. This

[7] https://ibm.github.io/terramind/

makes it a valuable tool for both benchmarking frameworks like NeuCo-Bench and real-world scenarios where fast, meaningful compression or semantic indexing of satellite data is needed.

OPENEO- EMBEDDINGS

OpenEO is a community API standard that gives EO researchers the ability to connect to backends offering EO processing capabilities in the cloud. Such unified access to EO data and processing resources enables researchers to run their analysis scripts on multiple backends with ease, fostering reproducibility of analysis results. Backend implementations exist in various languages, among them Python, Scala, Java and R (Figure 4).

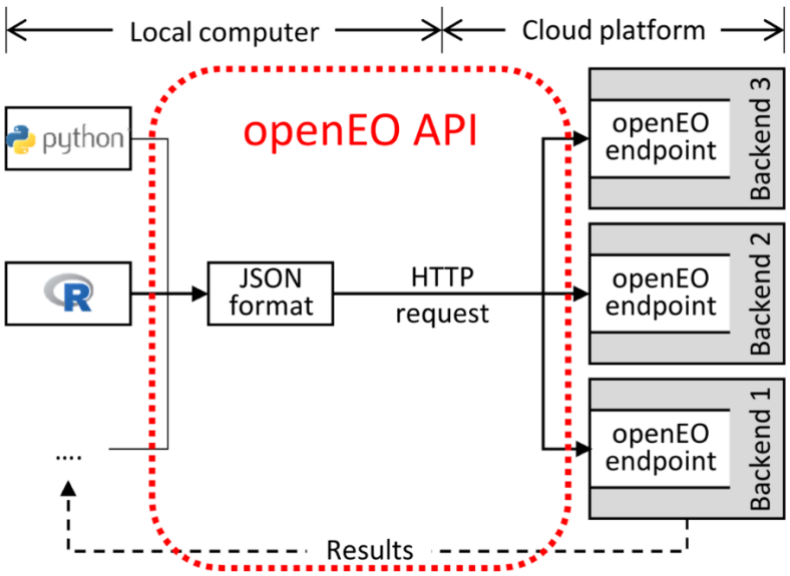


Figure 4: Architecture of the OpenEO API and endpoints.

OpenEO is supported by several major EO data providers (e.g., Euro Data Cube, VITO, Sentinel Hub). In Embed2Scale partners are extending openEO to handle embeddings and to enable embedding federation.

EMBEDDING-ENERGY-MONITOR

The Jülich Supercomputing Centre is developing EMBEDDING-ENERGY-MONITOR, a tool designed to quantify the energy consumption and associated CO_2 emissions required to generate embeddings from foundation models in Earth observation. By instrumenting inference pipelines with real-time hardware-level energy tracking, the tool estimates per-sample or per-batch energy usage across different models, hardware configurations, and input sizes. This enables researchers to estimate the environmental cost of producing embeddings, a crucial but often overlooked aspect of foundation model deployment.

The aim of EMBEDDING-ENERGY-MONITOR is to bring transparency and accountability to the resource footprint of foundation model inference. Its insights can inform model design choices, hardware selection, and workflow optimization strategies, guiding the community toward more sustainable and energy-efficient AI practices. In the long term, it could support green AI reporting

standards, enable carbon-aware benchmarking (e.g., in NeuCo-Bench), and empower users to make informed trade-offs between accuracy, speed, and environmental impact.

VESSEL-DETECTOR

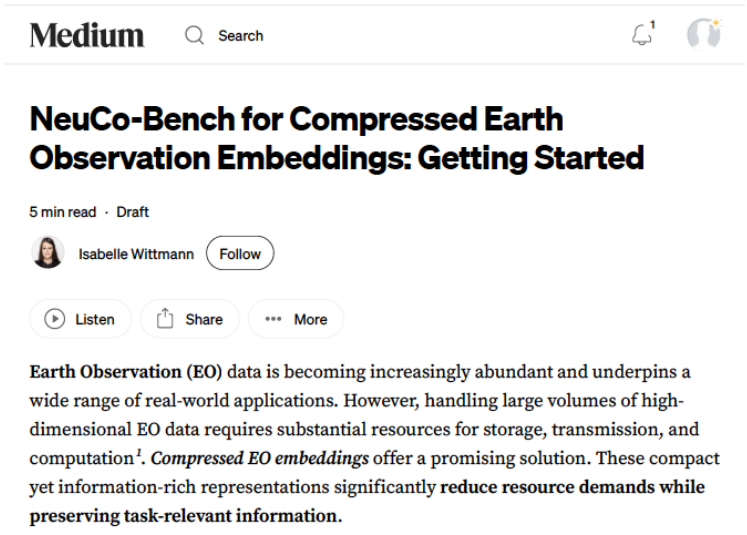
As part of the Embed2Scale initiative, vessel detector models have been developed to automatically identify ships in satellite imagery over the Mediterranean Sea. These models leverage foundation model embeddings and lightweight classifiers to efficiently detect vessels across large spatial and temporal scales. By focusing on high-resolution optical and synthetic aperture radar (SAR) data, the system can operate under various weather and lighting conditions, enabling near-continuous monitoring of maritime activity.

The ability to automatically detect vessels is crucial for applications such as border security, illegal fishing prevention, maritime traffic monitoring, and humanitarian response. The Mediterranean is a region of high geopolitical and environmental sensitivity, where timely and reliable information about vessel movements can support international cooperation and enforcement efforts. Embed2Scale's scalable detection pipeline demonstrates how embedding-based methods can transform raw EO data into actionable insights for operational maritime surveillance.

EMBEDDING WORKSHOP MATERIAL

As part of the project, we developed internal and external workshop materials (Figure 5) focused on the use of compressed embeddings in Earth Observation (EO). These materials guide participants through designing custom neural compressors and understanding embedding-size trade-offs using tools like NeuCo-Bench.

The workshops also demonstrate how to apply compressed embeddings in EO applications using models and libraries such as TerraTorch and openEO, showcasing tasks Embed2Transfer, Embed2Infer, Embed2Train and Embed2Find. Aimed at developers and environmental scientists, the materials promote efficient, scalable, and sustainable use of EO data.


Medium Search

NeuCo-Bench for Compressed Earth Observation Embeddings: Getting Started

5 min read · Draft

Isabelle Wittmann Follow

Listen Share More

Earth Observation (EO) data is becoming increasingly abundant and underpins a wide range of real-world applications. However, handling large volumes of high-dimensional EO data requires substantial resources for storage, transmission, and computation[1]. ***Compressed EO embeddings*** offer a promising solution. These compact yet information-rich representations significantly **reduce resource demands while preserving task-relevant information.**

Figure 5: Example of a Medium Blog as external workshop material.

4 STAKEHOLDERS

Embed2Scale serves a broad range of stakeholders across the Earth observation, AI, satellite, and policy ecosystems by delivering scalable, efficient, and sustainable solutions for embedding and processing large-scale geospatial data. The project fosters cross-sector collaboration, bringing together industry, academia, public authorities, and standardization bodies to collectively advance the use of AI and HPC in EO applications. By offering open tools, benchmarks, and best practices, Embed2Scale supports the growing need for intelligent, low-footprint geospatial analysis across Europe and beyond. The beneficiaries and stakeholders of Embed2Scale assets are:

1. Researchers, academia, and Copernicus users benefit by directly adopting Embed2Scale technologies and consuming compressed embeddings. Academic researchers and developers can use the provided tools (e.g. TerraTorch Embeddings, openEO for Embeddings) and benchmarks (e.g. NeuCo-Bench, Climate-Bench- Press) to explore new methods for efficient representation learning and downstream analysis. In turn, this engagement enriches the Embed2Scale ecosystem with cutting-edge insights and practical use cases.

2. Satellite data providers, as well as industry stakeholders, gain from improved interoperability and performance in their data pipelines. By leveraging the Embed2Scale framework and interlinked infrastructures (e.g. openEO embedding federation), these players can better align their services with user demands while adopting state-of-the-art embedding-driven solutions. Participation through Earth2vec enables knowledge exchange and ensures that SMEs and tech innovators can compete and scale efficiently in the EO data economy.

3. Public authorities and policy makers use Embed2Scale models (e.g. Vessel Detection) to inform strategic decisions around satellite infrastructure, digital investments, and environmental monitoring priorities and boarder control. The project supports initiatives from ESA around advanced data dissemination through embeddings from Copernicus Data Space Ecosystem. This helps ensure that public investments align with novel technological capabilities.

4. Standardization and open source bodies, such as OGC, will benefit from the best practices and architectural principles developed within Embed2Scale. The project offers guidance on embedding sustainability, interoperability, and efficiency into satellite network design and EO data handling. By contributing to community standards through Eart2Vec and reference implementations of openEO, Embed2Scale accelerates the broader adoption of emdedding driven EO systems across multiple domains.

5 MARKET SIZE

The Geospatial Imagery Analytics Market is rapidly expanding, projected to grow from USD 12.12 billion in 2025 to USD 18.55 billion by 2030, at a CAGR of 8.9%, according to MarketsandMarket[8]. A major driver of this growth is the increasing demand for advanced analytics tools capable of converting vast volumes of satellite and aerial imagery into actionable intelligence. Satellite imagery is expected to maintain the largest market share by data source, owing to its wide coverage, high frequency, and growing affordability due to commercial satellite proliferation. These data streams power diverse applications, including environmental monitoring, urban planning, agriculture, and defense sectors that require scalable, timely, and precise insights.

Industries are increasingly relying on geospatial analytics not only for core operations but also for strategic decision-making and risk management. For example, the insurance sector is poised for the fastest growth in geospatial applications, as high-resolution imagery improves risk assessment and claims validation, especially in the face of more frequent climate-related disasters. Similarly, sectors like agriculture, logistics, and infrastructure are adopting geospatial tools to optimize resource allocation, monitor asset conditions, and plan for future growth. As more industries move toward real-time, data-driven decision-making, the integration of AI, machine learning, and cloud computing with geospatial imagery analytics becomes essential for both operational efficiency and competitive advantage.

Embed2Scale assets directly address the growing demand from industry, government, and academic stakeholders for more efficient and scalable Earth Observation data access and processing. Traditional approaches rely on transferring and storing large volumes of raw satellite imagery, which are costly, slow, and computationally intensive. Embed2Scale enables the shift toward compressed feature representations (embeddings), computed server-side using foundation models trained with compression objectives. By retrieving these compact, semantically rich vectors instead of raw data, users benefit from reduced egress and storage costs, lower latency, and lighter computational loads for both training and inference tasks.

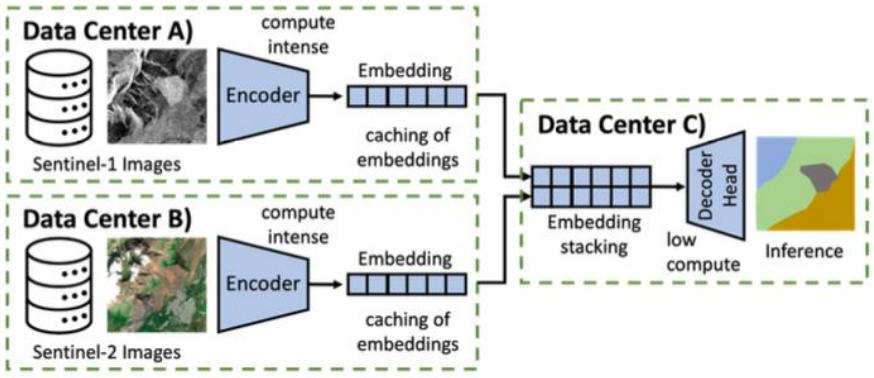


Figure 6: Efficient multi-modal EO analytics basdd on embedding federation.

[8] https://www.marketsandmarkets.com/Market-Reports/geospatial-imagery-analytics-market-221633264.html?utm_source=prnewswire.com&utm_medium=paidpr&utm_campaign=geospatial-imagery-analytics-market

This approach is particularly impactful for organizations with limited compute resources or those requiring low-latency, high-frequency analytics. Instead of downloading full-resolution imagery, stakeholders can receive pre-computed embeddings that are ready for downstream use, removing the need to train large backbone models. Real-time applications and global-scale analyses, such as deforestation tracking or agricultural monitoring, become significantly more accessible. The use of embeddings also facilitates federated data fusion, allowing users to combine datasets stored across different clouds (Figure 6), and supports fast similarity search in large EO archives using vector databases. Overall, Embed2Scale empowers a broader user base to leverage EO data through scalable, efficient, and intelligent infrastructure.

6 COMMUNITY SHAPING EARTH2VEC COMMUNITY

openEO Consoritum

Members of the Embed2Scale consortium actively participate in the openEO consortium (Figure 7), contributing to monthly coordination calls and technical discussions. As part of this collaboration, they are leading the development of an openEO API extension that enables the exchange, management, and federation of compressed embeddings across distributed EO infrastructures. This extension facilitates standardized access to embedding representations generated by foundation models, allowing users to query, retrieve, and combine features from multiple backends. The goal is to enhance interoperability and reduce the complexity of working with large-scale EO data, making it easier for developers and scientists to integrate embeddings into analytical workflows using the openEO ecosystem.

Earth2Vec Community

The Embed2Scale consortium launched the Earth2Vec community (Figure 8) at the ESA Foundation Model event in Frascati. The Earth2Vec community is an open, collaborative initiative focused on learning and sharing means to develop large-scale embeddings from Earth observation (EO) data. It brings together researchers, developers, and institutions to define best-practice, exchange information, to promote open benchmarks and develop community standards. The community emphasizes transparency, reproducibility, and sustainability, encouraging contributions of models, datasets, and tools.
We currently run monthly Earth2Vec calls, established a slack channel and exchange documents through box. The global community consists of 44 members and includes major players distributing embeddings like Asterisk Labs, Clay Foundation, Development Seed, LGND, Earth Genome, CloudFerro, ESA Phi-Labs, DLR, IBM, JFZ, Planet Labs, OpenGeoHub, Element84, Satlantis, Clark University, Google and Degas.

Figure 7: Earth2Vec community logo.

Figure 8: Earth2Vec community logo.

7 STANDARDIZATION

The Embed2Scale initiative is actively working to establish community standards for both the development and exchange of compressed Earth Observation (EO) embeddings through two complementary efforts:

EMBEDDING DEVELOPMENT STANDARDS IN THE EARTH2VEC COMMUNITY

Within the Earth2Vec community, we are addressing open challenges in how embeddings are created, structured, and stored. Key questions remain unresolved, such as how to optimally tile input data, how to manage pre-computation and caching, and how to balance compression vs. flexibility across spatial, spectral, and temporal dimensions.

For instance: Large EO tiles (e.g., Sentinel-2's 110×110 km) are often too coarse for neural network input and fail to capture fine-grained variability. Smaller tiles improve granularity but increase storage and compute costs. Pre-computing embeddings can boost performance but may limit user flexibility if not aligned with flexible tiling or streaming strategies.

These technical issues underscore the need for shared best practices and consensus on representation choices, which Earth2Vec is addressing through monthly community calls, shared documentation, and active collaboration among over 40 global partners.

EMBEDDING EXCHANGE STANDARDS IN THE OPENEO ECOSYSTEM

On the interoperability and access side, Embed2Scale members are extending the openEO API to support the exchange, management, and federation of compressed embeddings across distributed EO infrastructures. This includes:

- Defining open formats for embedding storage and transfer,
- Supporting metadata standards that capture model provenance, spatial/temporal coverage, and compression details,
- Ensuring compatibility with FAIR data principles by promoting discoverability through catalogs like STAC (SpatioTemporal Asset Catalogs).

Currently, no established standard exists for storing or exchanging neurally compressed EO embeddings, and platforms like Google Earth Engine and openEO do not yet support them natively. As such, our work contributes to filling this gap, with the goal of enabling consistent, machine-actionable exchange of embeddings, facilitating downstream analytics and broader adoption.

8 PARTNER EXPLOITATION

In addition to the general exploitation strategy, all consortium partners have developed and will update their own plans to improve the opportunities for exploitation of the Embed2Scale assets according to their role in the EO data and analytics value chain (Figure 9). In the final exploitation deliverable (D5.4) partners will also provide more details on the business model and incl. metrics like ROI and provide a risk analysis on external dependencies like the availability of platforms as Copernicus Data Space Ecosystem. The consortium will also consider to establish a Joint Exploitation Agreement until the end of the project.

Figure 9: Landscape of Embed2Scale consortium partners and advisory board members in terms of their position in the EO data and analytics value chain.

Current exploitation plans of the consortium are presented below:

SINERGISE

Assets to exploit: TerraTorch- Embeddings; Embedding-Energy-Monitor; openEO- embeddings; Earth2Vec Community, Embedding Workshop Material

Sinergise, now part of Planet, is a main Earth observation (EO) data provider and developer of Copernius data hub. They will use TerraTorch- Embeddings to generate embeddings on their proprietary data. They will also benefit from the Embedding Energy Monitor to reduce energy cost. Further, they aim at influencing or integrating openEO- Embeddings API on Copernics Data Space Ecosystem.

Further, they will benefit from Embedding Workshop Material to educate their internal workforce on the latest advances in AI in the EO domain. They will also participate in the Earth2Vec community calls, to stay at the forefront of foundation model development.

IBM

Assets to exploit: NeuCo-Bench, Compressor- Models; TerraTorch- Embeddings; Embedding-Energy-Monitor; openEO- embeddings; Earth2Vec Community

IBM will access the market for Embed2Scale assets through the Business Unit named IBM Environmental Intelligence[9]. Embed2Infer and Embed2Train will be key, to unlock the full potential of IBMs Geospatial Studio for near real-time regional assessments and rapid-fine tuning of AI applications.

IBM Research will continue to develop novel Compressor-Models and thus, will benefit from NeuCo-Bench and Embedding Energy-Monitor. Further, they will support CloudFerro to generate compressed embeddings on Copernicus Data Space Ecosystem to demonstrate the benefits of embeddings. Further, IBM Research aims at federating embeddings through openEO with DLR.

Romeo Kienzler and Isabelle Wittmann will continue to lead the Earth2Vec community by monthly calls, conference sessions and meetings, to establish a community standard for embeddings.

FZJ

Assets to exploit: Compressor- Models; TerraTorch- Embeddings; Embedding-Energy-Monitor; openEO- embeddings; Embedding Workshop Material; Earth2Vec Community

Forschungszentrum Jülich (FZJ) is evolving into a leading AI Factory, harnessing its cutting-edge computing infrastructure and research expertise to drive innovation in artificial intelligence. With its strong foundation in high-performance computing and data science, FZJ is well-positioned to offer advanced Earth Observation (EO) analytics services. By combining AI with EO data, FZJ can support applications in environmental monitoring, climate research, and sustainable development, enabling faster, more accurate insights for science, industry, and public decision-making.

In this context of the AI Factory, FZJ will deploy and operate Compressor- Models, use the Embedding-Energy-Monitor to minimize the carbon footprint and use TerraTorch- Embeddings to generate a catalogue of embeddings, and serve them through the openEO- embeddings API. Initiatives like DestinationEarth, EU research partners like JRC and ESA will benefit from the integration.

Beyond the technical Embed2Scale assets, FZJ will benefit from Embedding Workshop Material for the HDCRS summer school and other workshops at conferences like IGARSS.

The HDCRS Summer School is an annual event organized by the IEEE GRSS HDCRS Working Group, focused on advanced computing methods for Earth observation (EO). It brings together students, researchers, and professionals for lectures and hands-on training in cutting-edge EO technologies. FZJ plays a key role by contributing expertise in high-performance and AI-driven computing, helping participants apply these technologies to real-world EO challenges.

[9] https://www.ibm.com/products/environmental-intelligence

DLR

Assets to exploit: NeuCo-Bench, TerraTorch- Embeddings; openEO- embeddings; Earth2Vec Community, Embedding Workshop Material; Earth2Vec Community

The German Aerospace Center (CLD) will take advantage of NeuCo-Bench to test compressors to be used on-board of satellites, reducing data rates to ground stations. They are closely interacting with advisory board members ICEYE and Orora Technologies on this topic.

Further, they will use TerraTorch- Embeddings, to compute embeddings on TerraByte and make them accessible through the openEO- Embeddings API extension. The aim is to demonstrate embedding federation and its performance with IBM and the CDSE implementation of Sinergise.

The TerraByte[10] cluster at DLR is a high-performance computing infrastructure dedicated to large-scale Earth observation (EO) analytics and machine learning. It supports the training and deployment of foundation models on petabyte-scale EO datasets, providing researchers and institutions with the computational resources needed for embedding generation, model benchmarking, and data-intensive experimentation. TerraByte serves both internal DLR projects and external partners, including openEO communities, enabling collaborative development and evaluation of next-generation EO analytics.

The Embedding Workshop Material will be used for internal education of the workforce at the DLR institute for Software under the umbrella of the KnowledgeInterchangeWorkshops (WAW) which are a series of interdisciplinary workshops across DLR's organisational units with the aim of long-term knowledge exchange on interdisciplinary topics and the expansion and creation of professional networks.

SATCEN AND HISDESAT

Assets to exploit: Vessel-Detector; TerraTorch- Embeddings; Embedding Workshop Material

SATCEN will deploy the Vessel-Detector in production for analysts to perform automatic vessel detection for internal use and governmental stakeholders from EU member states. The model will be used for near real-time observations across the entire Mediterranean Sea, reducing labour cost from manual annotations and improve service performance.

Knowledge gained in Embed2Scale will be further used, to develop additional models. Therefore, internal workshops and with stakeholders from the EU member states will take advantage of the Embedding Workshop Material.

SATCEN aims at using TerraTorch- Embeddings to improve the data transfer form Hisdesat's PAZ data and hence improve the service quality for their customers.

[10] https://docs.terrabyte.lrz.de/about/introduction/

UOXF

Assets to exploit: Climate-Bench- Press; Embedding Workshop Material

The University of Oxford will further develop, distribute, and apply Climate-Bench-Press as a standardized evaluation framework for key stakeholders, including climate model developers and operators. Institutions such as ECMWF (member of the E2S advisory board) which are actively involved in the design process to ensure alignment with operational needs. Climate-Bench-Press will also support and benefit major Horizon Europe projects of UOXF, including ESiWACE3 (grant no. 101093054) and NextGEMS (grant no. 101003470), by providing robust tools for benchmarking next-generation climate models. In 2026, they aim at extending the Climate-Bench- Press with a climate model evaluation package including several down-stream tasks, which will also be open-sourced.

The University of Oxford is integrating the Embedding Workshop Material into its curriculum at the Intelligent Earth Centre for Doctoral Training, a UK-based graduate school. The material is used to deliver a dedicated class on embedding methods and applications, providing hands-on training for pre-doctoral researchers. This course equips students with practical skills in working with compressed embeddings for Earth system science and climate applications.

WWU

Assets to exploit: openEO- embeddings; Earth2Vec Community

The University of Münster (WWU) actively benefits from Embed2Scale assets such as the openEO embedding extension and the collaborative Earth2Vec community to advance open, interoperable Earth observation (EO) research. By integrating embeddings into the openEO framework and aligning with catalogue community standards, WWU supports open science, reproducibility, and data interoperability, core principles of modern geospatial research.

Through participation in openEO community calls alongside partners like EODC, Eurac, and VITO, WWU contributes to shaping the future of embedding-based workflows and metadata practices.

The Earth2Vec community further enables WWU to promote a common format for storing, exchanging, and cataloguing embeddings, directly benefiting the EO scientific community.

These assets also form a foundational element of WWU's proposed contribution to the Nationale Forschungsdateninfrastruktur 4 Earth (NFDI4Earth), and are being used to initiate community engagement through conference workshops and sessions at events such as AGILE, IGARSS, and EGU.