Platforms, Digital Infrastructure & Interoperability

Wed. Nov 11, 2026 4:00 PM - 5:30 PM JST | Wed. Nov 11, 2026 7:00 AM - 8:30 AM UTC | Room C (201) (2F)

Oral B1 Digital Infrastructure, Interoperability and AI-Enabled Platforms

Chair: Jon Blower, Naoki Hirose

5:15 PM - 5:30 PM JST | 8:15 AM - 8:30 AM UTC

[eHvadpaV-06] A Provenance-Preserving LLM-Assisted Interface for Heterogeneous Marine Observations in Digital Twins of the Ocean

*Jaronchai Dilokkalayakul[1] (1. Graduate School of Information Sciences, Tohoku University (Japan))

Digital Twins of the Ocean require interfaces that can integrate heterogeneous observations, support analysis, and preserve the provenance of system outputs. This is essential because DTO results may combine direct observations with processed products or analyses generated on demand. Each result must therefore remain traceable to its source data and processing history.

Marine ecosystem monitoring draws on information collected at different scales. Satellite-derived sea surface temperature describes broad surface conditions, while in situ CTD profiles capture changes through the water column. DNA-based community profiles add biological information from individual sampling sites. These sources are difficult to use together because they follow different formats, resolutions, identifier schemes, and scientific conventions. We therefore developed a provenance-preserving retrieval-augmented generation framework for coastal ecosystem monitoring.

Source-specific pipelines standardize key fields without discarding original metadata or provenance records. Related observations are linked through stable monitoring-event identifiers and represented as traceable evidence units in a shared knowledge base. The system retrieves evidence using both lexical and vector search, then incorporates relevant precomputed analyses and reliability assurance. A locally deployed large language model uses this context to produce grounded answers with citations. It also indicates whether information is observed or derived and identifies where evidence is missing.

We evaluated seven system configurations through an ablation study to examine the contribution of individual components. Precomputed analyses improved the completeness of answers, particularly when reliability information was included. The main remaining limitation was incomplete retrieval when questions required evidence from multiple sources. The framework offers an auditable interface for connecting marine observations with analytical functions and user inquiries within Digital Twins of the Ocean. Future work will activate on-demand analysis and assess the system with marine ecosystem researchers.

© 2026 DITTO Summit Local Organising Committee. All rights reserved.