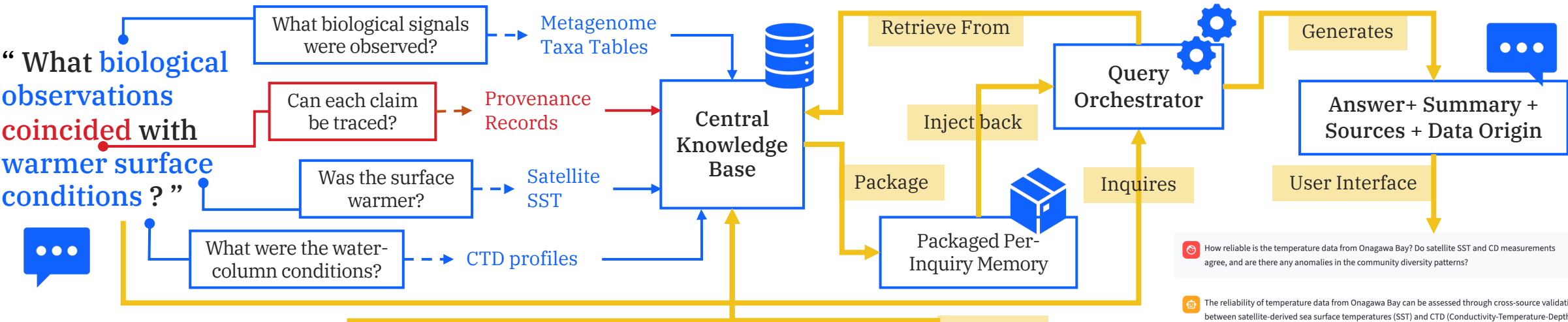


# Provenance-Preserving Framework for Integrating Heterogeneous Ecosystem Monitoring Data with Large Language Models

Jaronchai “Potter” Dilokkalayakul | Personal Site: [www.jaronchai.com](http://www.jaronchai.com) | Email: [dilokkalayakul.jaronchai.p8 \[at\] dc.tohoku.ac.jp](mailto:dilokkalayakul.jaronchai.p8[at]dc.tohoku.ac.jp)



## Data Sources

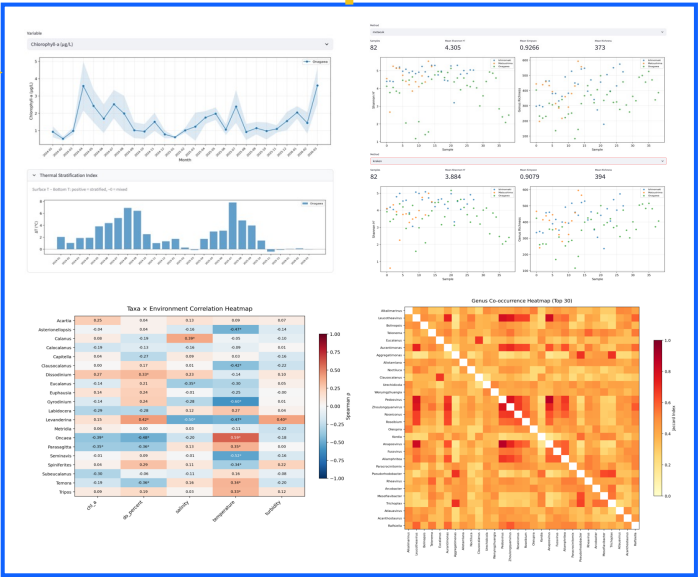


| Source                 | Key Attributes                                                                                  | Integration                                         |
|------------------------|-------------------------------------------------------------------------------------------------|-----------------------------------------------------|
| CTD profiles           | Structured water-column observations<br>Temperature, salinity, depth, oxygen, related variables | Sampling date/time<br>Station or bay<br>Depth layer |
| Metagenome eDNA tables | Taxonomic abundance and sample metadata<br>Genus / species-level biological signal              | Sample ID<br>Sampling date<br>Site or bay           |
| Satellite SST Himawari | Gridded sea-surface temperature<br>Spatiotemporal raster / NetCDF subset                        | Timestamp<br>Latitude/longitude<br>Bay region       |

Pre-Processed → Insert

Pre-Compute

Structured ecological computations are precomputed before the user asks a question, analysis



How reliable is the temperature data from Onagawa Bay? Do satellite SST and CD measurements agree, and are there any anomalies in the community diversity patterns?

The reliability of temperature data from Onagawa Bay can be assessed through cross-source validation between satellite-derived sea surface temperatures (SST) and CTD (Conductivity-Temperature-Depth) measurements. According to [reliability\_sst\_ctd\_validation], there is a high level of agreement between these two sources, with all 24 paired observations agreeing within  $\pm 2.0^{\circ}\text{C}$ , indicating strong reliability.

The mean absolute difference in temperature ( $|\Delta T|$ ) between satellite SST and CTD surface temperatures for Onagawa Bay is  $0.92^{\circ}\text{C}$ , which suggests that the measurements are very close to each other. The mean reliability score of 0.771 further supports this high level of agreement.

Regarding anomalies in community diversity patterns, [reliability\_diversity\_prediction] indicates one significant anomaly where the predicted Shannon diversity index deviates from the actual observed value by more than 2 standard deviations ( $\sigma$ ). This occurred on 2024-07-0-s1 in Onagawa Bay, with a predicted  $H'$  of 3.453 and an actual value of 1.601 (-2.3 $\sigma$ ), indicating a substantial deviation from expected diversity levels based on environmental conditions.

In summary:

- The temperature data from Onagawa Bay is highly reliable due to the strong agreement between satellite SST and CTD measurements.
- There are no reported anomalies in the cross-source validation for temperature, but there is one significant anomaly noted in community diversity patterns.

All 15 sources feeding the LLM

Retrieved Documents (6)

[sst\_2026-01-10] Satellite SST 2026-01-10  
Score: 0.0159 | remote\_sensing | 2026-01-10

Satellite SST for the Onagawa region on 2026-01-10. Regional statistics: mean=10.96°C, min=2.06°C, max=13.33°C (from 23 hourly observations). At monitoring station (38.4291°N, 141.4776°E): daily mean SST = nan°C (24 measurements).