In-Situ and IoT Data in Context of Earth Observation Data

Bernard Pruin, Nils Junike, Petabite GmbH, email: forename.surname@petabite.eu

Introduction

We present a data service (https://ieo.to) with functionality to search and access earth observation remote-sensing data and in-situ or IoT data in a unified way to drive efficient data discovery, download and execution of processing algorithms.
A hierarchical metadata aggregation scheme provides an efficient identification of in-situ data of interest over a large scale of query intervals.
The service supports both the discovery process and the long-term storage aspect for the in-situ data by providing an optimized compressing data storage.

In-situ and remote sensing data

Remote sensing data is characterised by wider area coverage with a potentially high sampling rate due to the ever finer spatial resolution [2]. The granularity of the data products is driven by various factors like convention, existing reference systems, file size consideration, orbit characteristics or processing needs. In general, remote sensing data products tend to be stored in relatively large units.
In-situ data is measured by a large variety of different sensors with less consensus and convention as there is for remote sensing data. While some standardization efforts exist, e.g. through the OGC Sensor Things API [3], there are a variety of ways such data is stored and presented through a wide range of individual services and solutions, e.g. [1]. Common approaches include the provision of file stores with files per device, possibly fragmented by time, database stores with query capability, packages covering sensor groups by type or geolocation for defined time periods.
In-situ data in the context of remote sensing is important for calibration, validation and the development of new algorithms.

Define and submit

IEOTO comes with a few integrated in-situ data sets, however a key element of the service is the possibility to integrate user owned in-situ data. Data submitted by a user is only visible to the submitting user.

Define Network

In-situ data submission is organised in Networks. Within a network there is a configuration per attribute that allows to validate input and interpret the values for visualisation and subsequent analysis:

```
{ "name": "pm10",
  "unit": "µg/m³",
  "label": "PM10",
  "channelType": "FLOAT",
  "description": "Particulate matter of
                  diameter 10 µm or less" }
```

Data from this network are submitted to the service in JSON format:

```
{ "pm25": 28.6,
  "sensorType": "SDS011",
  "pm10": 30.5
}
```

Submit in-situ data

Data handling

To make in-situ data searchable in a similar way as earth observation data products, without enforcing the associated granularity constraints, we generate virtual data products that fit the query asked.
To achieve acceptable search durations in a large range of scenarios, we use an aggregation hierarchy. The hierarchy levels are the same for all data streams and are hour, day, month, year and ALL_TIME. For each aggregation level and each variable, statistical summary data is maintained.
Within IEOTO the data storage structures and metadata aggregation hierarchy for search are decoupled. Storage is optimized for compression and granularity while still maintaining a reasonable balance between data to read versus data to provide for download requests in most scenarios.

Data discovery

Discover in context

When searching is performed for EO and in-situ data simultaneously. An appropriate aggregation level relative to the requested time interval is chosen to limit the number of potential responses for in-situ data. The reported metadata is in nearly all cases an approximation of the true value due to the difference between the aggregation interval and the search interval. For more precise searches drilling down in time can be used.

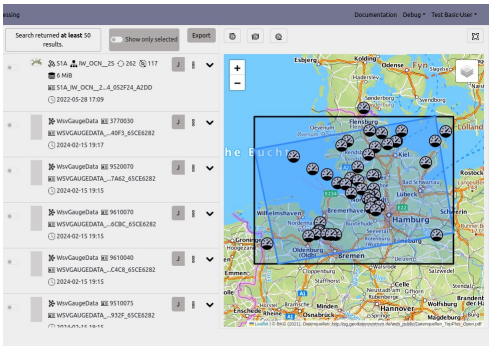
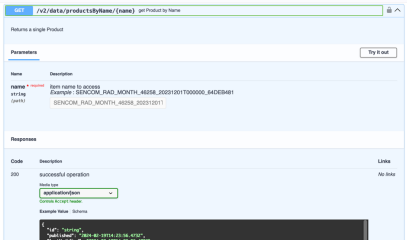


Figure : IEOTO combined search result EO product and in-situ data

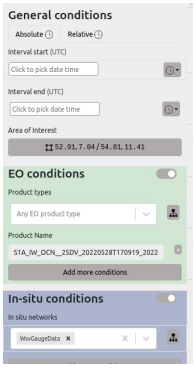
IEOTO provides a responsive web UI but can equally well be interfaced through a REST API that provides a superset of the UI functionality.



Download and use

Service data and functions

The service offers operationally the search for EO data products, AIS aggregated data products, hosted processing with in-situ data (AIS, ADS-B) that can be directly queried and retrieved as well as operational continuous storage of AIS, ADS-B and a collection of LoraWAN Sensors via the services sensor ingestion interface.



Currently available from Petabite

- Sentinel-1 GRDH ship sighting packages based on AIS data [4]
- Sentinel-2 MSIL2A ship sighting packages based on AIS data [4]
- Sentinel-1 and Sentinel-2 AIS data in product realm [4]
- Citizen science air quality data
- Selected AIS ship data since 2020
- Selected ADS-B flight data since 2022

https://ieo.to

References

[1] ESA: https://scihub.copernicus.eu/userguide/, retrieved 05/2022.
[2] Copernicus in situ TAC: Product User Manual for multiparameter Copernicus In Situ TAC (PUM), V1.12, 2021, DOI: 10.13155/43494.
[3] OGC SensorThings API. https://docs.ogc.org/is/18-088/18-088.html. 2021-08-04.
[4] Petabite Data Service IEOTO. https://ieo.to/
[5] Petabite data type documentation. https://eitlab.com/petabite.eu/documentation/petabite-datatypes
[6] Petabite service API. https://eitlab.com/petabite.eu/documentation/pb-ieoto-public-api
[7] Loew, A., et al. (2017), Validation practices for satellite-based Earth observation data across communities, Rev. Geophys., 55, 779–817, DOI:10.1002/2017RG000562.
[8] Sensor.community web site. https://sensor.community/