In-Situ and IoT data inventory in support of earth observation data analysis

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

Introduction

We present a data service 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 search, download and execution of processing algorithm development and operational production.

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 value adding process and the long term storage aspect for the in-situ data by providing an optimized compressing data storage layer without compromising the efficient retrieval of data portions of interest.

Remote sensing

Remote sensing data usually covers wider areas with a potentially high sampling rate due to the ever finer spatial resolution. Setting the emerging video data products aside, as a convention the EO data is provided in the form of data product of limited spatio-temporal extent [2], where the granularity of the data products may be 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

For In-situ data that is measured by a large variety of different sensors no similar consensus has yet emerged. While some standardization efforts exist, e.g. through the OGC Sensor Things API, there are in reality 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. The latter approach comes closest to the concept of data products for remote sensing data.

In-situ data in the context of remote sensing is important for calibration, validation and the development of new algorithms.

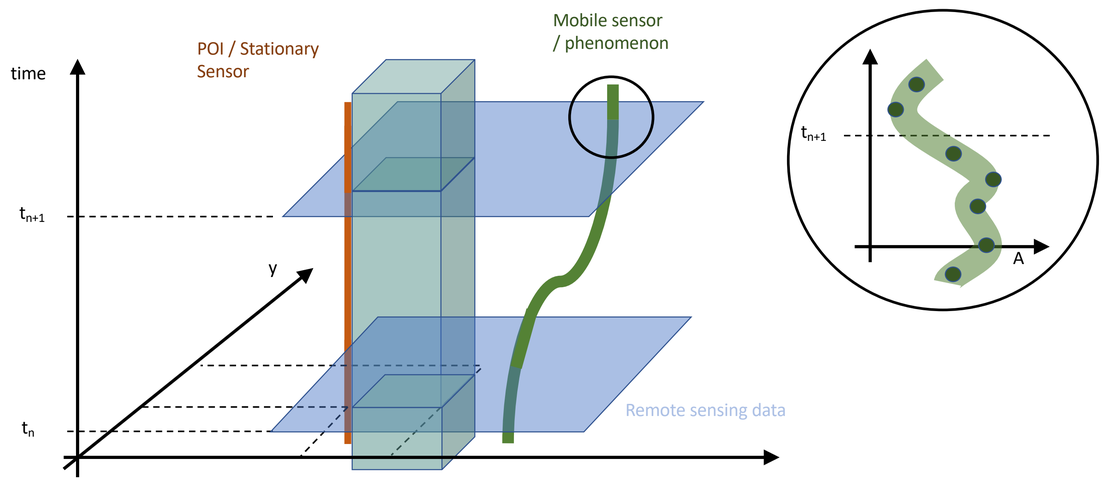

Figure: Remote sensing vs. in-situ data characteristics

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. The design goals of our service solutions are to assemble query results fast, independent off the in-situ data sampling frequency to provide meaningful metadata describing the data found in a meaningful way allow searching on metadata that represents the value range of the data within the search interval.

References

[1] Copernicus in situ TAC: Product User Manual for multiparameter Copernicus In Situ TAC (PUM), V1.12, 2021, DOI: 10.13155/43494.

[2] ESA: https://scihub.copernicus.eu/userguide/, Retrieved 05/2022.

[3] OGC SensorThings API. https://docs.ogc.org/is/18-088/18-088.html . 2021-08-04.

[4] Petabite GmbH: Petabite Data Types, https://gitlab.com/petabite.eu/documentation/petabite-datatypes, Retrieved 05/2022

[5] Petabite Data Scout service. https://petabite.space/

Data handling

To achieve acceptable search durations in a large range of scenarios, we use an aggregation hierarchy that limits the number of search results in relation to the query interval. The hierarchy levels are the same for all data streams and are 10 min, 1h, 1d, 1m, 1y, complete. For each aggregation level and each variable, maxium, minium, average as well as time of first and last measurement in interval are kept and aggregated. When searching, an appropriate aggregation level relative to the requested time interval is chosen to limit the number of potential responses. We have opted for a pragmatic approach with respect to the resulting metadata. The resulting 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 a reduction in the search interval is required.

Within the service, there is no direct relationship between the data storage structures and metadata aggregation hierarchy for search. 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.

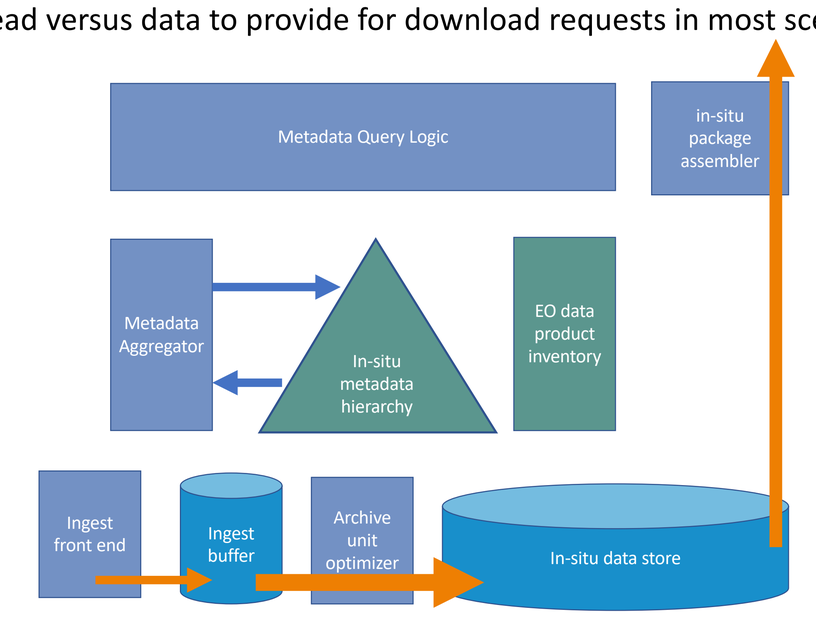

Figure: EO and in-situ data combined search infrastructure concept

Special cases

The service functionality is not optimal for some limiting cases. Moving sensors that travel long distances are not well captured by aggregation. For ships and planes (in the future) the service therefore provide pre-computed dedicated data structures that explicitly relate to applicable earth observation data products.

When there are a large number of sensors involved in the search area, the result presentation per sensor becomes untenable. For this special casea we are in the process of developing a result presentation that aggregates further by sensor type.

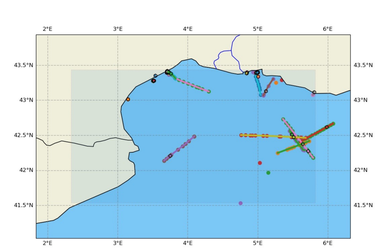

Figure: Pre-computes ship tracks in relation to existing EO product

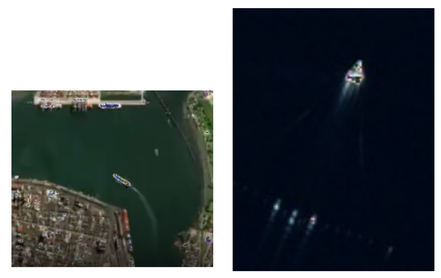

Figure: Pre-processed EO products containing AIS-tracked ship traces

Service functions

The service currently 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 Test Sensors via the services sensor ingestion interface. Users may also store their own in-situ data as input for their hosted processes.

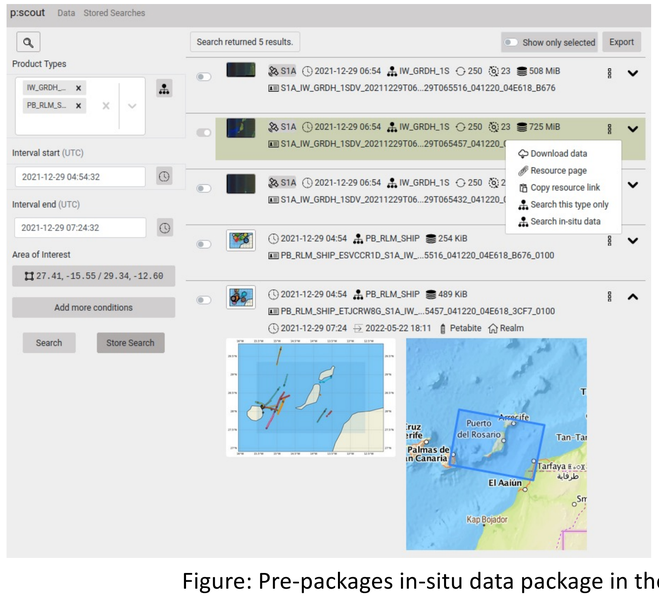

Figure: Pre-packages in-situ data package in the Petabite data service

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]