Data Management Plans

1. Data Collection

Collecting marine traffic data from Finnish Transport Infrastructure Agency’s data sources.

* Harbor Schedules gathered from the Portnet-system
* Vessel location AIS (Automatic Identification System)
* Dirways from Baltice-system
* Sea state estimation data from TLSC-system, that analyzes data send by smart AtoN buoy sites
* Waterway traffic disturbances

1. Data Preprocessing

Process the data through ETL process

* Extract

Identify the output of API code (JSON, CSV) and then save API response to JSON if the API return JSON data or CSV if need. After that load into pandas for analysis.

* Transform

Clean the data, Handling outliers etc.

* Load

Store the processed data in postgresDB.

1. Data Processing & Orchestration

Use Apache Airflow to schedule and automate data processing workflows

1. Machine Learning (ML) Integration

Implement ML algorithms to predict real-time vessel movements and other marine traffic patterns

1. Data Visualization

Create dashboards and visual representations of the processed and predicted data