Online Analytical Processing of Port Calls for Decision Support

Aidan Worth
Delevant Business Solutions Ltd
Nicosia, Cyprus
aidan.worth@delevant.com

Aris Televantos
Delevant Business Solutions Ltd
Nicosia, Cyprus
aris.televantos@delevant.com

Nicos Evmides
Cyprus University of Technology
Limassol, Cyprus
nicos.evmides@cut.ac.cy

Michalis P. Michaelides
Cyprus University of Technology
Limassol, Cyprus
michalis.michaelides@cut.ac.cy

Herodotos Herodotou
Cyprus University of Technology
Limassol, Cyprus
herodotos.herodotou@cut.ac.cy

***Abstract*—The port call process encapsulates a visitation cycle of a ship to a port and can generate a wealth of data. The real time analysis of port call data can be used to find bottlenecks in the port call process, establish targets based on key performance indicators (KPIs), and to understand how shipping traffic impacts a port's efficiency. This demonstration will showcase a new Power BI interactive report powered by a multidimensional OLAP cube for very fast performance, which is built on top of a data warehouse collecting information from various sources in real time. The report currently visualizes several KPIs and other types of information that can be filtered per port, time-period, vessel type, origin or destination ports, and various other categories to help manage arrivals, departures, and port operations.**

***Index Terms*—port calls analysis, maritime decision making**

I. Introduction

The port call process entails the whole procedure between the ship arrival to, and departure from a particular port. This requires the collaboration of many different actors (e.g., ship, pilots, tugboats, terminals, port authority) each with their unique demands and aims. Thus, the collaboration and cooperation of all actors are very important, especially when it comes to sharing data in real-time and being ready for just-in-time operations to facilitate this process. The Port Collaborative Decision Making (PortCDM) concept (and platform) aims to increase the digital collaboration among actors involved in a port call [1]. Its main purpose is to record timestamps for the states a vessel completes during its visitation cycle to a port. PortCDM captures estimated/actual timestamps associated to vessel arrivals/departures to/from various locations such as port area, berth, anchorage, etc., and beginning/ending times of various operations such as cargo operations, anchoring, etc.

The data collected in the PortCDM platform encompass a wealth of information that can be analyzed to provide meaningful insights and support decision-making activities for all involved actors [2]. Traditionally, this kind of analysis is done while a data snapshot is taken, transformed, and cleansed, and then fed into one-time calculations and analysis that provide results of specific Key Performance Indicators (KPIs) through tables, reports, charts, and other visual aids [3]. When new data becomes available, the whole process has to be repeated, involving quite a few manual steps, until a new set of reports is generated. This inflexibility leads to unanswered questions and missed opportunities to act preemptively [4].

The key goal of this work is to make this process of data extraction, transformation, load, and analysis of port call data, live and seamless as to allow for a continuous analysis and interpretation of the events at a port. Such analysis can be used to find bottlenecks, establish targets, and to understand how shipping traffic patterns impact a port's efficiency. Further benefits of online analytical processing include (i) optimizing the allocation of incoming vessels to berths and improving berth occupancy; (ii) building a comprehensive set of internal KPIs that help every level of operations understand their targets; and (iii) measuring attainment of these KPIs in a continuous fashion to show ships that fell outside targets.

The backend technology used to support this effort is based on a curated data warehouse and an On-Line Analytical Processing (OLAP) cube, which consumes data from the data warehouse. This OLAP cube is then utilized by a Power BI report for secure high-performance reporting via a user's web browser. This demo will showcase the Power BI report containing two dashboards that produce actionable information to support executive decisions. These dashboards are fully interactive with KPIs that can be filtered or 'sliced and diced' per port, time-period, vessel type, ports, etc. In one dashboard, the user can interact with the various graphs to explore and extract information regarding port call count, predictability, and punctuality regarding arrivals, departures, and berthing times. In the other dashboard, significant productivity KPIs are visualized based on various filtering conditions.

II. Port Calls Data Mining Framework

Data mining is the process used to extract actionable information from a larger set of raw data. Raw data can be in

This work was co-funded by the European Regional Development Fund and the Republic of Cyprus through the Research and Innovation Foundation (STEAM Project: INTEGRATED/0916/0063).

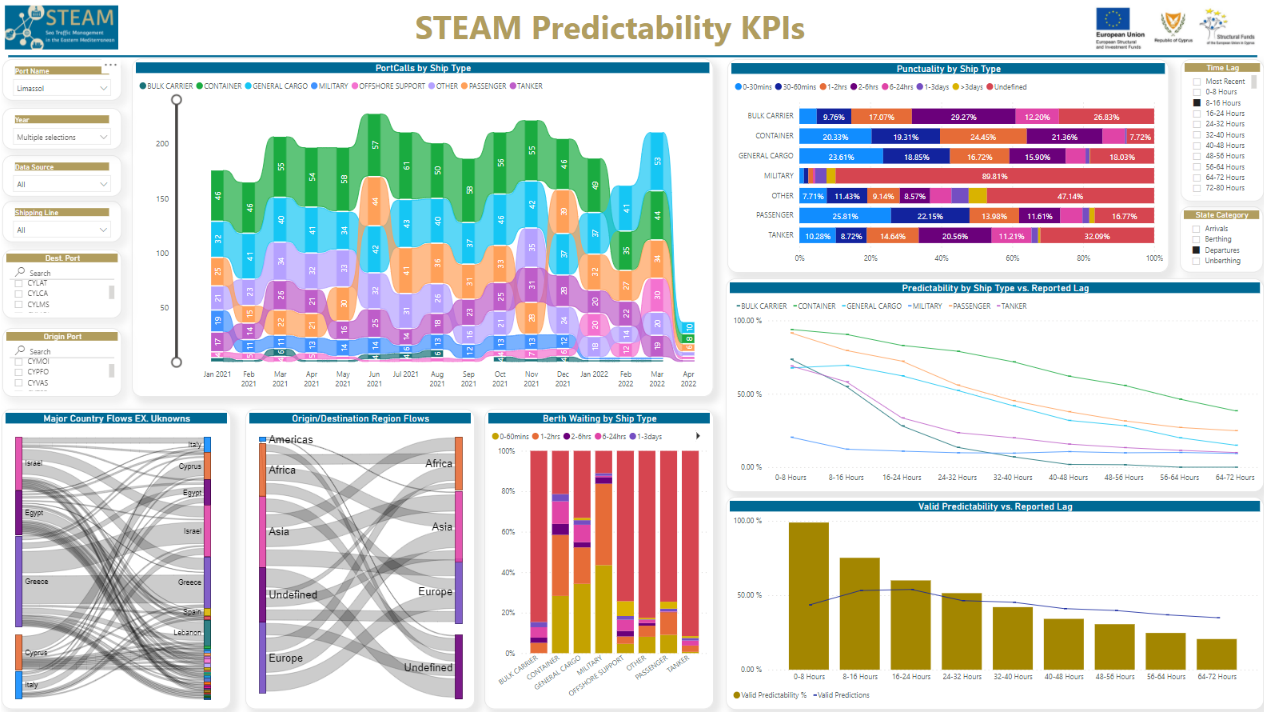

Fig. 1. Power BI dashboard with predictability KPIs.

a format that is very difficult or time intensive to data mine or requires the combination/pre calculation/intelligent filtering in several steps to facilitate the data mining process. This preparation work is referred to as Extract, Transform, and Load (ETL), which must be an automated and robust operation that combines several different data sources into a Data Warehouse that is structured to provide calculated data in a format that is favorable to the later data mining procedures. The entire process consists of multiple steps described below.

Data Collection: The original data is stored in different relational databases. Combining these separate sources of information into a report can be difficult because they typically do not share a unified reporting data schema. A mixture of custom T-SQL and CodeGen is used to create, maintain, and link the data into a common data warehouse.

Data Cleaning: Cleaning and standardizing the data aims for unified reporting across various different data systems and sources of data. For this purpose, we make use of Fuzzy matching, which is a general technique for finding strings that match a pattern approximately rather than exactly. To perform this we utilize Microsoft's SQL Server Integration Services (SSIS) and T-SQL's adaptive CodeGen.

Data analysis: The product of data collection and cleaning is a data warehouse that contains a single entry for each action of the vessel in the port call process such as an actual/estimated time of arrival/berthing/departure, etc. Data is not aggregated in the data warehouse but in views used to produce various reports and metrics. The data warehouse is not the final data layer for the production of dashboards as it does not offer the features needed for optimum reporting ability (high number of simultaneous queries and concurrent users). Instead, an *OLAP Cube* is built, sourcing information directly from the data warehouse. An OLAP Cube is a data structure that allows fast analysis of data according to multiple dimensions that define a business problem. They are better suited for creating records from a series of transactions known as On-Line Transaction Processing (OLTP) and fast end-user interaction with data.

III. Key Performance Indicators

Dashboards contain charts that cover a select list of Key Performance Indicators (KPIs) and can be used to evaluate port efficiency in relation to the following metrics.

Predictability is the degree to which a correct prediction or forecast of a state (e.g., vessel arrival at the port, vessel departure from the port, berthing, etc.) can be made.

$$Predictability = 1 - \frac{|AT - ET|}{|AT - Reported(ET)|} \tag{1}$$

AT denotes actual time, ET estimated time, and $Reported(ET)$ the time when the estimate was reported. Predictability can vary depending on when the prediction is made. It is apparent that if the reported time of a prediction is closer to the actual event, better predictions can be made. But sometimes, decisions need to be made at a specified time before the event (e.g., 2 days before the ship arrives). A prediction of arrival after this point can be useless or should not be accounted for. The dashboards were developed to take into account this factor, i.e., when the prediction was made, and this is called a *time lag*. For example, a lag of 24 hours

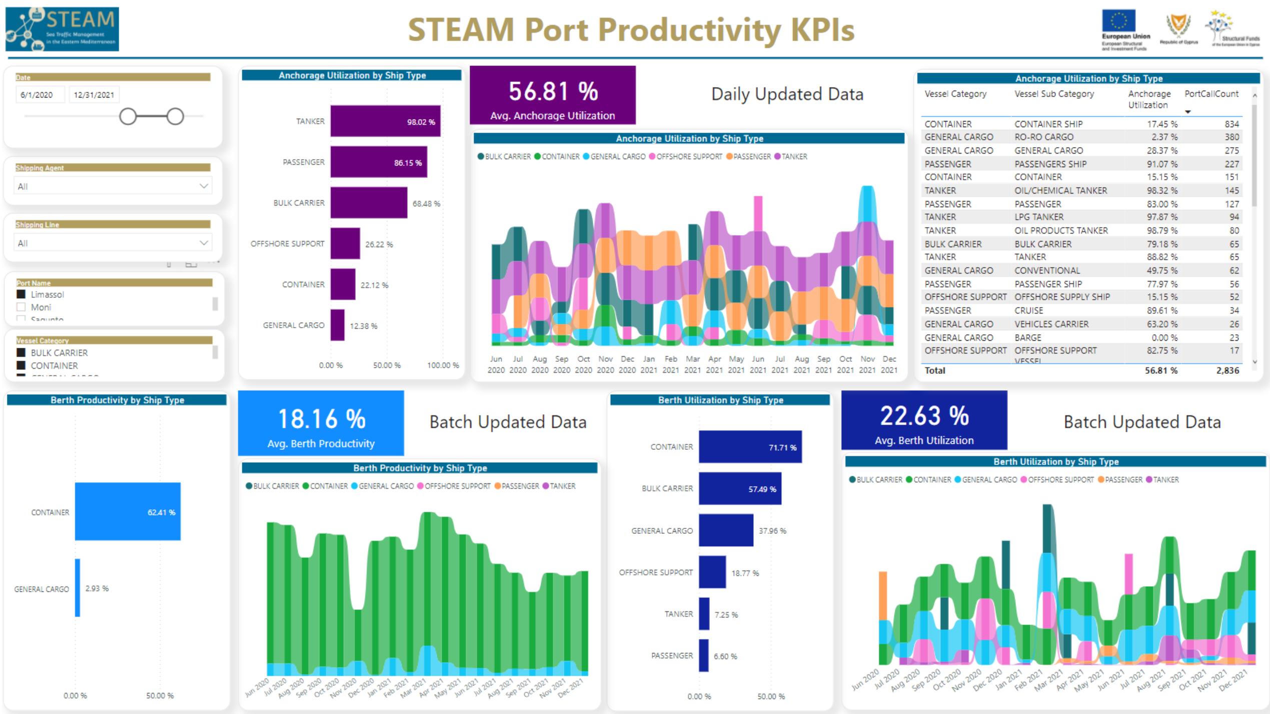

Fig. 2. Power BI dashboard with productivity KPIs.

means that we are counting predictability for predictions made 24 hours or more before the actual event. In many cases, the correct lag for accurate and helpful analytics is a matter of debate. Hence, the dashboard exposes information for all lags, letting the user select the predictability KPIs for a specific lag depending on the situation.

Punctuality is the percentage of vessels arriving, departing, berthing, or unberthing within a period (e.g., 30 minutes) of the estimated time to do so:

$$Punctuality = \frac{Port\ calls\ s.t.\ |AT - ET| < 30\ min}{Count\ all\ port\ calls} \times 100 \quad (2)$$

Berth Waiting is the percentage of vessels waiting to berth for some period (e.g., 60 min) after arriving at the port:

$$Berth\ Waiting = \frac{Port\ calls\ s.t.\ |ATB - ATA| < 60\ min}{Count\ all\ port\ calls} \times 100 \quad (3)$$

Berth Utilization for a port call is the percentage of time spent at berth divided by the total time spent in the port:

$$Berth\ Utilization = \frac{ATUB - ATB}{ATD - ATA} \times 100 \quad (4)$$

Berth Productivity for a port call is the percentage of time spent performing cargo operation at berth divided by the total time spent at berth; hence, it depicts the efficiency of a berth visit related to the purpose of call:

$$Berth\ Productivity = \frac{AT\_COp\_Fin - AT\_COp\_Start}{ATUB - ATB} \times 100 \quad (5)$$

IV. Data Analytics Dashboard

The current report provides a detailed breakdown of the predictability and productivity KPIs in an interactive fashion that allows for cross filtering by port, year, data source, shipping line, destination port and origin port.

The first dashboard shown in Figure 1 displays (from top left, clockwise): (1) Port call count, analyzed by ship type and month; (2) Punctuality by ship type, segmented by time lag (difference of reported ET from AT); (3) Predictability by ship type segmented by time lag; (4) Valid predictability vs reported time lag segmented by time lag; (5) Berth waiting by ship type, segmented by duration; (6) Ship flows by origin/destination region; (7) Ship flows by top-5 origin/destination countries.

The second dashboard shown in Figure 2 displays (from top left, clockwise): (1) Total anchorage utilization by ship type;(2) Anchorage utilization by ship type segmented by month; (3) Anchorage utilization by ship type with vessel sub category and port call count; (4) Berth utilization by ship type segmented by time; (5) Total berth utilization by ship type;(6) Berth productivity by ship type segmented by month; (7) Total berth productivity by ship type.

References

[1] M. Lind *et al.*, "Port Collaborative Decision Making–Closing the Loop in Sea Traffic Management," in *14th COMPIT*, 2015.

[2] M. P. Michaelides *et al.*, "Port-2-port Communication Enhancing Short Sea Shipping Performance: The Case Study of Cyprus and the Eastern Mediterranean," *Sustainability*, vol. 11, no. 7, p. 1912, 2019.

[3] N. Andrienko and G. Andrienko, "Visual Analytics of Movement: An Overview of Methods, Tools and Procedures," *Information visualization*, vol. 12, no. 1, pp. 3–24, 2013.

[4] M. Michaelides *et al.*, "Decision Support in Short Sea Shipping," in *Maritime Informatics*. Springer, 2021, pp. 221–236.